HANDBOOK *of* THE HISTORY OF LOGIC

VOLUME 5

LOGIC FROM RUSSELL TO CHURCH

Edited by
Dov M. Gabbay
John Woods

NORTH-HOLLAND

Handbook of the History of Logic

Volume 5
Logic from Russell to Church

Edited by

Dov M. Gabbay
Department of Computer Science
King's College London
Strand, London, WC2R 2LS, UK

and

John Woods
Philosophy Department
University of British Columbia
Vancouver, BC Canada, V6T 1Z1
and
Department of Computer Science
King's College London
Strand, London, WC2R 2LS, UK
and
Department of Philosophy
University of Lethbridge
Lethbridge, Alberta, Canada, T1K 3M4



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
North-Holland is an imprint of Elsevier



North-Holland is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
Linacre House, Jordan Hill, Oxford OX2 8DP, UK

First edition 2009

Copyright © 2009 Elsevier B.V. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at http://elsevier.com/locate/permissions, and selecting *Obtaining permission to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-444-51620-6

For information on all North-Holland publications visit our website at elsevierdirect.com

Printed in Hungary

08 09 10 11 10 9 8 7 6 5 4 3 2 1

PREFACE

In designing the *Handbook of the History of Logic*, we have been guided by two quite general principles of organization. In some cases a volume would show to advantage if organized around schools of thought or theoretical types. *Logic and the Modalities in the Twentieth Century* and *The Many Valued and Nonmonotonic Turn in Logic* are examples of this principle at work. In other cases we have found it more natural to assemble volumes around the names of practitioners. The subtitle of *The Rise of Mathematical Logic: From Leibniz to Frege*, reflects this principle, as does the present volume, *Logic From Russell to Church.* To a considerable extent, the two organizing principles are a loose convenience, between which the lines of demarcation are somewhat arbitrary. Even so, the developments tracked in this present volume are, by and large, at least as well known by the names of their authors as by their own names.

We have been asked for our rationale in ending volume 3 with Frege and opening volume 5 with Russell. Were not Frege and Russell virtual collaborators in giving to logic the momentum that would drive its mathematical turn? Why, then, would this pair be treated separately in the *Handbook*? Part of the answer to this question is unapologetically utilitarian. Volumes can only be so big, and chapters can be published only when they are ready. But the answer also has a further part, concerning which a word or two of explanation would be in order. It is well known that Frege and Russell shared a philosophical motivation for producing a new logic. The motivation is logicism, which asserts the reducibility of arithmetic to pure logic. Notwithstanding this commonality, Frege and Russell had quite different positions about the epistemic status of logic. Frege thought that the axioms of logic were known with certainty. Russell's view was that in some cases an axiom — reducibility, for example - could be justified only inductively (or, as we would say, abductively) or pragmatically. It would not be too much to say that this difference about mathematical epistemology presages the instrumentalism that seeks to mediate the rivalrous pluralisms that would be exhibited by logic in the years ahead. A further difference between the two logicists pertains to the background context of their work. Frege did most of his work on mathematical logic before he knew of the paradox communicated to him by Russell in that fateful letter of 1902. Frege's adjustment to it was an afterthought published at the page-proof stage as an appendix to volume two of the *Grundgesetze.* For Russell on the other hand, virtually all of his contributions to logic were fashioned in the aftermath of his paradox. His work was deeply marked by efforts to evade it, and was so in ways that gave further impetus to the instrumental turn in logic, albeit

one that Russell tried to resist in the name of "logical common sense". But it would not be too long before Carnap would proclaim his Principle of Tolerance:

> "In logic there are no morals. Everyone is at liberty to build up his own logic ... as he wishes."

We may say, then, that the short period between the publication of the second volume of the *Grundgesetze* (1903) and of *Principles of Mathematics* (1903) marks the beginning of the end for a naïve apriority about the principles of logic, an epistemological shift that would reconfigure the explosive intellectual successes of the first fifty years of the twentieth century.

Logic's two dominant themes in the period from Russell to Church are its foundational creativity and its procedural limitations and liabilities. One might say that this was a period accentuated by both triumph and crisis. This is the time when logic would acquire its modern face, what with the emergence of model theory, proof theory, recursion theory, set theory and metamathematics. Hence the great names: Russell, Brouwer, Heyting, Skolem, Löwenheim, Herbrand, Wittgenstein, Ramsey, Lesniewski, Hilbert, Gödel, Tarski, Post, Gentzen, Curry and Church. But it was also the time in which these foundational breakthroughs would have to be engineered in the face of, in addition to Russell's paradox, the semantic paradoxes of Tarski and others and the incompleteness results of Gödel, all of which which tested Hilbert's insistence that mathematical truth has no home beyond the safe-harbours of formal proof. The problems that attended the emergence and development of mathematical logic were not only difficult; they were also treacherous. The first great technical accomplishment in logic was Aristotle's all-but-correct proof of the perfectibility of syllogisms. There are those who think that the mediaeval doctrine of *suppositio* ranks high, as it surely does in the form in which categorical propositions are construed as multiply-quantified statements of identity. The 20^{th} century originators of modern logic faced crises that threatened the collapse of two of its foundations — set theory and model theory — and a third which significantly attenuated the scope of a further foundation — proof theory. The great ambitions of logic would have to be navigated beyond the reach of these perils. Meeting these difficulties demanded, and received, system-building of extraordinary virtuosity, unparalled by anything that had gone before. From this there emerged a stable idea of what it took to be a logic: a precise, and typically a recursively generated, grammar; a well-developed formal semantics; a syntactic theory of formal proof; and the capacity to answer its own metatheoretical questions - soundness, completeness, decidability and the like. In the fifty-year period covered by the present volume, mathematization became a necessary part of what it took to be a logic; in the following fifty years many would see mathematical virtuosity as sufficient. For the most part the founders tried to honour Russell's principle of logical common sense. While they sought to produce mathematically robust systems, they also wanted these systems to be conceptually sound. The paradoxes were a wrench to such ambitions. It is to the everlasting credit of the founders that the paradoxes were no bar to mathematical virtuosity

(indeed they were in large part the *provocateurs* of it). But the same could not be said of conceptual fidelity. The founders knew this, and said so; the age of intuitive foundations for logic was coming to an end. What for the founders was a philosophical discouragement has been for at least some of their descendents a naïve and dispensable presumption. Concerning the concepts that logicians have an interest in analyzing, what has come to count above all is the mathematical adequacy of the systems that describes them. Conceptual adequacy has become a separate issue at best.

The founders of twentieth-century logic were mainly philosophers, and those who weren't were respectful of philosophical motivations. Their investigations of logic were driven by philosophical considerations and judged by the standards of philosophical soundness, such as they may be. But as the fortunes of mathematical virtuosity rose and those of philosophical cogency diminished, a corresponding transformation was underway in the administrative arrangements under which intellectual work was institutionalized and professionalized. What began as a philosophical discipline became, in the space of five decades, a mathematical one. The irony is unmistakable. In the hands of Frege and Russell, the role of logic was a take-over of mathematics. Take-overs are tricky, whether the Norman occupation of Britain of Daimler's absorption of Chrysler. There is always the lurking possibility of "going native", which is precisely what happened to these two invaders. How could it also not happen to any logic that was purpose-built to capture arithmetic? In its take-over of mathematics, logic went native.

Some will think talk of "take-over" and "going native" too melodramatic for serious scholarship. But this is a stylistic objection, not a substantive one. The transformation of logic into mathematics is one of the momentous events of intellectual history. It is a transformation initiated by the very people who would not have wanted to see it happen. In remarking upon it here, it is not necessary to lament philosophy's diminution; one might just as well celebrate logic's liberation. Was it not a philosopher who remarked that logic is an old discipline, and that since 1879 is a great one? Even so, the question remains whether philosophy has a future in the developments in logic that lie ahead.

We should also say a word or two about the inclusion here of Meinongian logic. Meinong was not a logician and never thought of himself as one. So Meinong did not make our list of pantheonic names. Even so, Meinong had an elaborate theory of objects and a concomitant, if largely implicit, theory of reference for them. One of the most distinctive features of Russell's logic — and of most of those that followed in its wake — is the theory of reference reflected in the doctrine of logically proper names and definite descriptions. In some ways, Russell's theory of reference was conceived of as an alternative to Meinong's theory of objects. If we here allow alternativeness to be symmetric, then Meinong's implicit theory of reference is also an alternative to Russell's, which Russell took seriously enough to try to rebut, and it is interesting if only on that account. In what has come to be called "Meinongian logic", Meinong's theory of reference has been transformed into a formal semantics, one that pivots on a domain of discourse that reflects

the main features of Meinong's theory of objects. While Meinongian logic has not entered the 20^{th} century mainstream, it is the first modern example of a logic of existence-neutral quantification which, in the second half of the twentieth century, did become a contending logic for the treatment of negative existentials and so-called empty referring expressions.

Once again we are deeply and most gratefully in the debt of the volume's superb authors. We also warmly thank the following persons: Professor Margaret Schabas, Head of the Philosophy Department and Professor Nancy Gallini, Dean of the Faculty of Arts, at the University of British Columbia; Professor Michael Stingl, Chair of the Philosophy Department and Christopher Nicol, Dean of the Faculty of Arts and Science, at the University of Lethbridge; Jane Spurr, Publications Administrator in London; Carol Woods, Production Associate in Vancouver, and our colleagues at Elsevier, Senior Acquisitions Editor, Lauren Schultz and Gavin Becker, Development Editor.

Dov M. Gabbay
King's College London

John Woods
University of British Columbia
and
King's College London
and
University of Lethbridge

CONTRIBUTORS

Mark van Atten
University of Paris, France.
mark.vanatten@univ-paris1.fr

Serge Autexier
DFKI, Saarbrücken, Germany.
serge@dfki.de

Christoph Benzmüller
Saarland University, Saarbrücken, Germany.
chris@ags.uni-sb.de

Andrea Cantini
Università di Firenze, Italy.
cantini@philos.unifi.it

Felice Cardone
University di Milano-Bicocca, Milano, Italy.
cardone@disco.unimib.it

Jens Erik Fenstad
University of Oslo, Norway.
jfenstad@math.uio.no

J. Roger Hindley
University of Swansea, UK.
J.R.Hindley@swansea.ac.uk

Andrew D. Irvine
University of British Columbia, Vancouver, Canada.
a.irvine@.ubc.ca

Dale Jacquette
University of Bern, Switzerland.
Dale.jacquette@philo.unibe.ch

Juliette Kennedy
University of Helsinki, Finland.
juliette.kennedy@helsinki.fi

Joan Rand Moschovakis
(Occidental College, USA) and MPLA, Greece.
joan.rand@gmail.com

Jan von Plato
University of Helsinki, Finland.
jan.vonplato@helsinki.fi

Michael Potter
Cambridge University, UK.
michael.potter@phil.cam.ac.uk

Jonathan P. Seldin
University of Lethbridge, Canada.
jonathan.seldin@uleth.ca

Wilfried Sieg
Carnegie Mellon University, USA.
sieg@cmu.edu

Jörg Siekmann
DFKI, Saarbrücken, Germany.
siekmann@dfki.de

Keith Simmons
University of North Carolina at Chapel Hill, USA.
ksimmons@email.unc.edu

Peter M. Simons
University of Leeds, UK.
P.M.Simons@leeds.ac.uk

Hartley Slater
University of Western Australia, Australia.
slaterbh@arts.uwa.edu.au

Alasdair Urquhart
University of Toronto, Canada.
urquhart@cs.toronto.edu

Claus-Peter Wirth
Saarland University, Germany.
drcp@surfeu.de

BERTRAND RUSSELL'S LOGIC

Andrew D. Irvine

Bertrand Russell is generally recognized as one of the most important English-speaking philosophers, logicians and essayists of the twentieth century. Often cited along with G.E. Moore as one of the founders of modern analytic philosophy and along with Kurt Gödel as one of the most influential logicians of his time, Russell is also widely recognized for his sustained public contributions to many of the most controversial social, political and educational issues of his day. Even so, more than anything else, it is Russell's work in logic and the foundations of mathematics that serves as his core contribution to intellectual history and that makes Russell the seminal thinker he is. His most significant achievements include

1. his refining and popularizing of Giuseppe Peano's and Gottlob Frege's first attempts at developing a modern mathematical logic,

2. his discovery of the paradox that bears his name,

3. his introduction of the theory of types (his way of avoiding the paradox),

4. his defense of logicism, the view that mathematics is in some important sense reducible to logic, and his many detailed derivations supporting this view,

5. his ground-breaking advances in technical philosophy, including both his theory of definite descriptions and his theory of logical constructions,

6. his theory of logical relations, including his impressively general theory of relation arithmetic,

7. his formalization of the reals,

8. his theory of logical atomism, and

9. his championing of the many connections between modern logic, mathematics, science, and knowledge in general.

Russell's first important book about logic, *The Principles of Mathematics*, appeared in 1903. This was followed by the landmark articles "On Denoting" (in 1905) and "Mathematical Logic as based on the Theory of Types" (in 1908). A decade of technical work then culminated with the publication (in 1910, 1912 and

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

1913) of the three-volume *Principia Mathematica*, co-authored with Russell's former teacher, Alfred North Whitehead.[1] This was followed by several books and articles intended to help underscore the philosophical importance of the new mathematical logic that Russell, Whitehead, Peano and Frege had developed, including "The Philosophy of Logical Atomism" (in 1918–19), *Introduction to Mathematical Philosophy* (in 1919), "Logical Atomism" (in 1924), and *My Philosophical Development* (in 1959). Russell received the Order of Merit in 1949 and the Nobel Prize in 1950. Born in 1872, he wrote over fifty books, over 4,000 articles, book chapters, reviews and other more minor publications, and over 61,000 letters. He died in 1970, having lived a very full and public life.

Developing a comprehensive, coherent interpretation of Russell's logic and philosophy of mathematics is both important and challenging. It is important because of the tremendous influence his work has had on philosophy, mathematics, economics, decision theory, computability theory, computer science and other disciplines over the past one hundred years. It is challenging for at least three reasons. First, modern logic is intrinsically abstract and complex. Even mastering the full notation of *Principia Mathematica* may be judged an accomplishment in itself. (This is especially true given that, over the past century, many parts of the notation have been modified or replaced by more modern kinds of symbolism. Learning it in complete detail is thus something that not even all Russell scholars have achieved.) Second, as with the work of any intellectual pioneer, it is often suggested, not only that Russell's views changed and developed over time, but that at important times they were also relatively inchoate. Third, as with the work of many other important historical figures, it sometimes can be difficult to distinguish between Russell's actual views and those attributed to him by various well-known commentators and interpreters.

Today there exist two main approaches to understanding the most difficult and demanding aspects of Russell's philosophy. The first is to claim that Russell's work is largely a series of almost unconnected views, views that not only changed over time but that also were largely indebted to the work of others. This is the view that characterizes Russell primarily as a popularizer and as a less-than-systematic thinker. The second is to see in Russell's writings a much more systematic and original body of work. To give just one example, there are many often-overlooked connections between Russell's logic, his theory of language (especially his theory of definite descriptions), and his epistemology (for example, his multiple-relation theory of judgment). According to this second view we fail to see these many connections at our peril. Significantly, it is also this second view — that Russell is offering us a fully integrated body of intellectual work — that is much more in line with Russell's own statements about his philosophical aims and achievements, statements that many writers over the years have conveniently chosen to ignore.

[1] Although anticipated in several respects in the writings of Gottlob Frege, *Principia Mathematica* remains just as intellectually impressive an achievement as Frege's *Grundgesetze der Arithmetik*. The case might also be made that, without it, Frege's work might never have received the prominence it deserves.

Whether explicitly or not, it is also this integrated view of logical, philosophical and scientific knowledge that has influenced an entire century of western philosophy.

Understanding not only Russell's substantive logical and philosophical views, but also his methodological views, is therefore crucial to understanding the dominant paradigm of intellectual work carried out in philosophy over the past century. Today it is hard to imagine a philosophy undergraduate anywhere in the western world whose education has not been influenced, at least in part, by Russell's mathematical logic and philosophical method.

INITIAL INFLUENCES

> At the age of eleven, I began Euclid, with my brother as my tutor. This was one of the great events of my life, as dazzling as first love. I had not imagined there was anything so delicious in the world. After I had learned the fifth proposition, my brother told me that it was generally considered difficult, but I had found no difficulty whatever. This was the first time it had dawned upon me that I might have some intelligence. From that moment until I met Whitehead and I finished *Principia Mathematica*, when I was thirty-eight, mathematics was my chief interest and my chief source of happiness. Like all happiness, however, it was not unalloyed. I had been told that Euclid proved things, and was much disappointed that he started with axioms. At first I refused to accept them unless my brother could offer me some reason for doing so, but he said: "If you don't accept them we cannot go on," and as I wished to go on, I reluctantly admitted them *pro tem*. The doubt as to the premises of mathematics which I felt at that moment remained with me, and determined the course of my subsequent work. [Russell, 1967, pp. 37–8]

So begins the story in Russell's *Autobiography* of how Russell came to be interested in the foundations of mathematics. Educated at home by a series of Swiss and German governesses and tutors before entering Cambridge in 1890, Russell embraced academic life enthusiastically. He obtained a first in Mathematics in 1893. Even so, with the exception of the lectures given by Whitehead, Russell found the teaching of mathematics at Cambridge "definitely bad."[2] As a result, he became even more determined to try to learn exactly what it is that gives mathematics its high degree of certainty.

Having finished his undergraduate program in mathematics, Russell "plunged with whole-hearted delight into the fantastic world of philosophy"[3] and the following year he completed the Moral Sciences Tripos. In philosophy he was initially influenced by the absolute idealism of J. M. E. McTaggart, F. H. Bradley and the

[2] [Russell, 1959, p. 38].
[3] [Russell, 1959, p. 38].

other Ox-bridge neo-Hegelians of the time.[4] As Russell understood it, absolute idealism depended crucially on the doctrine of internal relations, the doctrine that any relational fact (for example, that x is related to y) is really a fact about the natures of the related terms. According to this view, if x is greater than y, then being greater than y is a part of the nature of x. Object y is thus, in some significant sense, a part of x and x, similarly, is a part of y. Given the complexity of relations in the world, it turns out that all objects must be related to all other objects. Hence it follows that there exists only a single, all encompassing unity. Further, since if one is aware of x, x must also (on this view) be a part of one's mind, it follows that everything conceivable must be a part of consciousness.

Initially, Russell used this idealist framework in his approach to mathematics, for example in his *Essay on the Foundations of Geometry* (1897). However, he became disenchanted with idealism once he realized how incompatible the view was with his developing understanding of both mathematics and science. On Russell's emerging theory, geometrical points, for example, could be individuated only by their relations. But according to absolute idealism, relations in turn depended on the intrinsic natures of their individual relata. Since, for the idealist, any relation holding between two or more distinct objects must be present in (or grounded in) the nature of the objects being related, relations must be internal to objects. In other words, the unreality of relations was an essential consequence of idealist doctrine. In yet other words, relations can never have an external existence or reality. Russell's revolt against idealism thus hinged on the reality of relations, on the idea that relations are, in some importance sense, independent of the objects being related, that they are objective and real in just the same way as are particulars.

Of course Russell admits that for some relations, the idealist view may turn out to be plausible. If Abelard loves Héloïse, in all likelihood the relation of love will simply reduce to certain kinds of mental states belonging to Abelard. Even an atheist, says Russell, will take it to be a fact that some people love God. Yet if God does not exist, this means that the "relation" of love need not rely in any realist way upon the second of the relation's two relata. Similarly, if Ponce de Leon desires to drink from the fountain of youth, and if there in fact is no such fountain, the "relation" of desire cannot plausibly be thought to be something holding between two separate and independent objects. Instead, the relation again will be understood most plausibly as simply a certain kind of state of mind of the individual having the desire in question. In this way, it will not require reference to either a (non-existent) fountain of youth or an external relation.

In contrast, in the case of many other relations, the facts seem to point towards an alternative theory. If e_1 and e_2 are events, and if e_1 is earlier than e_2, then it seems that there is nothing about e_1, and nothing about e_2, that determines that such a relation must in fact hold. Similarly, to cite another of Russell's favorite examples (due originally to Leibniz), if a man living in Europe has a wife in India who dies without his knowing it, it is nonsense to think that at the

[4] [Griffin, 1991].

moment of her death the man will undergo some form of intrinsic change. The relation between the two has been altered tragically and permanently, but the man, until he learns of the event, will remain unaffected. Since asymmetrical ordering relations (for example, the relation of being a predecessor, or of being earlier than) are essential to both mathematics and science, Russell saw immediately the importance of abandoning a doctrine of purely internal relations.

Forced to choose between idealism and mathematics, Russell thus unhesitatingly chose mathematics, replacing idealism with realism and replacing the doctrine of internal relations with a new doctrine that he called "the doctrine of external relations," the view that relations, like objects, have a reality independent of the objects they relate, not in the sense that they exist alone in the world, but simply in the sense that they are not equivalent to, or reducible to, any property or properties of the objects being related. On Russell's view, relations are thus something independent of (although not separate from) the individuals they relate. They are also (for the most part) independent of both mind and language.

This change in Russell's view also meant abandoning the idealistic doctrine of monism. Since the doctrine of internal relations holds that every relation is grounded in the natures of its related terms, and since every object is related to every other object, it follows that the universe is not a collection of separate, independent individuals but that, in some important sense, it is a single, all-encompassing unity. To reject the doctrine of internal relations thus meant accepting the three main factors underlying Russell's defense of the "new realism," namely

1. pluralism, the view that the universe is populated by numerous logically independent particulars and relations,

2. anti-psychologism, the view that mind and the world are independent, in the sense that the world (including both its logical and material components) does not depend upon mind for its existence, and

3. the scientific outlook, the view that advances in knowledge (including advances even in logical knowledge) occur only through careful analysis, the weighing of evidence, and the attitude that all opinions must be held tentatively, or with an awareness that new evidence may at any moment lead to their abandonment.

All three of these influences soon came to inform Russell's understanding of, and his expansion of, the new "mathematical logic" of his day.

MODERN MATHEMATICAL LOGIC

Russell's final break with idealism came in July 1900 when he met Peano. The International Congress of Philosophy had been organized that year to coincide with the Paris Exhibition, and it was in Paris that Russell finally discovered the building blocks he needed for his positive account of mathematical knowledge:

> The Congress was a turning point in my intellectual life, because I there met Peano. I already knew him by name and had seen some of his work, but had not taken the trouble to master his notation. In discussions at the Congress I observed that he was always more precise than anyone else, and that he invariably got the better of any argument upon which he embarked. As the days went by, I decided that this must be owing to his mathematical logic. I therefore got him to give me all his works, and as soon as the Congress was over I retired to Fernhurst to study quietly every word written by him and his disciples. It became clear to me that his notation afforded an instrument of logical analysis such as I had been seeking for years, and that by studying him I was acquiring a new and powerful technique for the work that I had long wanted to do. [Russell, 1967, pp. 217–18]

In particular, Russell found in Peano's work two advances that would prove essential to his purposes. The first was the introduction of the quantifiers. The second was the distinction between a unit class and its sole member. Together, these two advances allowed Russell (like Frege) to introduce the idea of a propositional function. Like more familiar mathematical functions, propositional functions are a kind of "incomplete symbol." By themselves they fail to express propositions. But once their variables are replaced with constants, the result in each case will be a proposition that is either true or false. Russell then used the idea of a propositional function to distinguish between names and definite descriptions and between classes defined by enumeration and classes defined by abstraction. An example helps make both of these conceptual developments clear.

Prior to the introduction of the quantifiers (i.e., the introduction of explicitly logical phrases such as "For all objects x", represented by the symbol "$\forall x$", and "For some objects x", represented by the symbol "$\exists x$"), the propositions "All Greeks are mortal" and "Socrates is mortal" were assumed to have much the same logical form. Both appeared to assign a predicate (such as "is mortal") to a subject (such as "Socrates"). In contrast, modern logic sees these two propositions as having quite different logical forms. "Socrates is mortal" still retains this subject-predicate form. Thus if M is introduced to stand for the predicate "is mortal", and s is introduced to stand for the individual Socrates, the proposition will be represented as Ms. In contrast, "All Greeks are mortal" takes the quite different form, "For all objects x, if x is Greek then x is mortal." If we allow G to stand for the predicate "is Greek" and M to stand for the predicate "is mortal," this form then can be represented in modern logical notation as "$\forall x(Gx \supset Mx)$". The result is that the second proposition is no longer seen as assigning a predicate to a subject. Instead, it is seen as asserting a relation between two propositional functions, Gx and Mx. Put in other words, despite initial appearances, the proposition "All Greeks are mortal" does not specifically talk about (or refer to) a given class (or set) of individuals, namely the Greeks. Instead it makes a (quite different) claim about every individual in the universe, saying that for any such individual, if it has the property of being a Greek, then it also has the property of being mortal.

Thus unlike "Socrates is mortal", the proposition "All Greeks are mortal" refers to no particular individual or class of individuals. Instead, it expresses a relationship between propositional functions (or predicates). For this reason it will be true (if in fact it is true), regardless of whether there are even any Greeks to which one may be intending to refer.

The distinction between a unit class (or unit set) and its sole member was equally important to Russell. As Russell saw it, the class of natural satellites of the Earth has only a single member, the Moon. But if we are to identify the two terms "natural satellite of the Earth" and "the Moon", the logic of classes (and hence the theory of numbers) becomes muddled in innumerable ways. In other words, because "the Moon" functions most often as a name, and so refers directly to the Moon regardless of whatever properties we may want to assign to it, it follows that it can be used successfully only by those who are aware of (or at least who are capable of making) this reference. In contrast, the term "natural satellite of the Earth" can be used successfully by all language users who understand the meaning of the phrase, regardless of whether they have direct acquaintance with the Moon, and regardless of whether the Moon even exists. Put in other words, by distinguishing between the unit class (whose sole member is the Moon) and the Moon itself, Russell was able to distinguish clearly between a referring term that names the Moon ("the Moon") and the property (being a natural satellite of the Earth) that defines the unit class having only this single member.

These two advances (the introduction of the quantifiers and the distinction between a unit class and its sole member) led Russell to a much clearer understanding of the idea of a propositional function. This, in turn, led him to a completely new way of understanding both logic and the foundations of mathematics.

RUSSELL'S PARADOX

Russell tells us that he finished the first draft of his *Principles of Mathematics* "on the last day of the nineteenth century — i.e. 31 December, 1900."[5] The months since meeting Peano had been the most fruitful and intellectually promising of his life. He had made tremendous advances, not only in understanding Peano's logic, but also in developing a formidable logic of relations. It was these developments, Russell believed, that would prove to be essential for his larger logicist ambition, that of proving mathematics to be reducible to logic. However, by May 1901, Russell's "intellectual honeymoon" came crashing to an end. As he tells us, "the honeymoon could not last, and early in the following year intellectual sorrow descended upon me in full measure."[6] The reason was the discovery of the paradox that now bears Russell's name.

Russell's paradox is the most famous of the logical or set-theoretical paradoxes. The paradox arises within naive set theory by considering the class of all classes

[5] [Russell, 1959, p. 73].
[6] [Russell, 1959, p. 73].

that are not members of themselves. Such a class appears to be a member of itself if and only if it is not a member of itself, clearly an impossible result.

To understand the paradox, it is best to begin with a simple example. Some classes, such as the class of all prime numbers, are not members of themselves. In other words, the class (or set or collection) of all prime numbers is not itself a prime number. Other classes, such as the class of all things that are not prime numbers, are members of themselves. In other words, the class or set or collection, of all things that are not prime numbers is itself not a prime number. Call the class of all classes that are not members of themselves "R." If R is a member of itself, then by definition it must not be a member of itself. Similarly, if R is not a member of itself, then by definition it must be a member of itself. Yet it must be one or the other and cannot be both. Discovered by Russell in the late spring of 1901, the paradox appeared to put an end to the new logic that Russell, Peano and Frege had been developing.

Cesare Burali-Forti, an assistant to Giuseppe Peano, had discovered a similar antinomy in 1897 when he noticed that since the class of ordinals is well ordered, it too must have an ordinal. However, this ordinal must be both an element of the class of all ordinals and yet larger than every such element. Believing that this puzzle likely resulted from a simple ambiguity in Burali-Forti's reasoning, Russell had not spent much time focusing on the problem. Instead, he came to his new paradox via a quite different route: by studying Georg Cantor's proof that there is no largest cardinal. Asking whether there is a largest class of objects led Russell to realize that some classes are members of themselves while others are not. This led to consideration of the class of all classes that are not members of themselves. This, in turn, led to the paradox. Unlike both Burali-Forti's paradox and Cantor's famous proof, Russell's paradox did not involve reasoning about either ordinals or cardinals and relied instead only on the primitive notion of a class. It thus struck Russell as almost purely logical in nature, and so that much more difficult to resolve.

The significance of the paradox can be seen once it is realized that, using classical logic, all sentences follow from a contradiction. For example, assuming both p and $\sim p$, any arbitrary proposition, q, can be proved as follows: from p we obtain $p \vee q$ by the rule of Addition; then from $p \vee q$ and $\sim p$ we obtain q by the rule of Disjunctive Syllogism. Because set theory underlies all branches of mathematics, Russell immediately began to worry that the inconsistency of set theory would show that no mathematical proof could be completely trustworthy. Far from having helped secure the foundations of mathematics, Russell had inadvertently discovered a new and devastating weakness.

Russell wrote to Frege with news of his paradox on June 16, 1902.[7] The paradox was of significance to Frege's work since, in effect, it showed that the axioms Frege was using to formalize his logic were inconsistent. Specifically, Frege's Rule V, which states that two classes are equal if and only if their corresponding functions coincide in values for all possible arguments, requires that an expression such

[7] Russell [1902]. Cf. [Frege, 1902].

as $f(x)$ be considered both a function of the argument x and a function of the argument f. In effect, it was this ambiguity that allowed Russell to construct R in such a way that it could both be and not be a member of itself.

Russell's letter to Frege arrived just as the second volume of Frege's *Grundgesetze der Arithmetik* (*The Basic Laws of Arithmetic*) was in press. Immediately appreciating the difficulty the paradox posed, Frege added to the *Grundgesetze* a hastily composed appendix discussing Russell's discovery. In the appendix Frege observes that the consequences of Russell's paradox are not immediately clear. For example, "Is it always permissible to speak of the extension of a concept, of a class? And if not, how do we recognize the exceptional cases? Can we always infer from the extension of one concept's coinciding with that of a second, that every object which falls under the first concept also falls under the second? These are the questions," Frege notes, "raised by Mr Russell's communication."[8] Because of these worries, Frege eventually felt forced to abandon many of his views about logic and mathematics.

THE THEORY OF TYPES

Like Frege, Russell was also concerned about the contradiction. Upon learning that Frege agreed with him about the significance of the discovery, he immediately began writing an appendix for his own soon-to-be-released *Principles of Mathematics*. Entitled "Appendix B: The Doctrine of Types," the appendix represents Russell's first significant attempt at providing a principled method for avoiding the contradiction.

Russell's paradox ultimately stems from the naïve idea that any condition (or property) may be used to determine a class. For example, the property of being evenly divisible only by itself and by one determines the class of prime numbers. The conjunctive property of being a vertebrate, being warm-blooded, having hair or fur, having an external ear, having lungs and a diaphragm, giving birth to live young and nourishing them with milk determines the class of mammals. The property of being both a square and not a square (or any other contradictory property) determines the empty class, and so on. As a result, Russell's mind immediately turned toward the possibility of restricting the principles governing class existence. Specifically, he asked whether there might be some way of restricting the so-called Abstraction (or Comprehension) axiom, the axiom that states that any propositional function, $f(x)$, containing x as a free variable can be used to determine a class. In other words, is it true that corresponding to every propositional function, $f(x)$, there will exist a class whose members are exactly those things, x, that have the property f?

In Russell's mind, any resolution to the paradox would have to meet three conditions. In descending order of importance these conditions were as follows: First, it would have to avoid the contradiction. Second, it would have to be

[8] [Frege, 1903, 127] (repr., p. 1).

compatible with retaining as much classical mathematics as possible. Third, it would have to be mathematically and philosophically plausible, and not simply an *ad hoc* solution. In other words, the solution would have to coincide, as much as possible, with what Russell called "logical common sense."[9]

In his appendix, Russell suggests that reference to troublesome classes (such as R) might be avoided by arranging all sentences into a hierarchy, beginning with sentences about individuals at the lowest level, sentences about classes of individuals at the next lowest level, sentences about classes of classes of individuals at the next lowest level, and so on. Then, using the so-called "vicious circle principle" (VCP) introduced by Henri Poincaré, Russell hoped to be able to explain why propositional functions, such as the function 'x is a class,' could not be applied to themselves: they could not be applied to themselves since self-application would involve a vicious circle. In other words, it is possible to refer to a class or collection of objects for which a given condition (or predicate) holds only if each member of the class is defined independently of (or prior to) the predicate in question. In Russell's terminology, it is possible to refer to a class or collection of objects only if they are all at the same level or of the same "type."

The reasoning behind the VCP may be summed up as follows: when we assert a proposition such as $\forall x(fx)$, if the assertion is to be significant, the values over which the variable x ranges will have to be well defined. (As Russell puts it, they must be *definite*.) In other words, if the function is to be unambiguous there will have to be a fixed totality of values over which x will range. Otherwise, the proposition will fail to have meaning. However, if we now proceed to create new values over which x will range, values that can only be defined in terms of that totality, the totality both will have to contain and not contain these new values. But if so, it follows that the totality cannot be definite or well defined. Of course we might attempt simply to include these new values in some new totality, but if they can only be created on the basis of a previous "totality," this attempt must always fail. In other words, the assertion will never "catch up" to itself, it will both be and not be a totality, it will never be capable of becoming definite.

The moral Russell draws from this argument is that propositions that refer to a given totality of propositions may never themselves become members of that totality. The result, if we want to be able to discuss this new type of proposition, will have to be to create a hierarchy of propositions. First-order propositions will turn out to be those that fail to refer to any totality of propositions; second-order propositions will turn out to be those that refer to totalities of first-order propositions; third-order propositions will turn out to be those that refer to totalities of second-order propositions; and so on *ad infinitum*. In this way, the vicious self-reference that Russell believed lay at the heart of Russell's paradox is eliminated.

Although first introduced by Russell in his appendix to the *Principles*, his theory of types eventually found its more mature expression in his 1908 article "Mathematical Logic as based on the Theory of Types" and in *Principia Mathematica*. This more mature theory involved two different kinds of hierarchy. First, there is a

[9] [Russell, 1959, p. 79].

simple extensional hierarchy consisting of individuals, classes of individuals, classes of classes of individuals, and so on. Second there is a hierarchy of propositional functions, consisting first of the so-called *predicative* functions (propositional functions which do not refer to any totality of functions), then of first-order functions (propositional functions which refer to the totality of predicative functions), then of second-order functions (propositional functions which refer to the totality of first-order functions), and so on. Thus, in its details, Russell's theory admits of two quite different versions, the "simple theory" and the "ramified theory." The need for the more complex theory apparently arose because, as Russell points out in his appendix, there is also a more insidious version of the paradox that depends upon the interaction between classes and propositions.

Because, in Russell's theory, propositions are objectively existing objects, it follows that they can be members of classes. But propositions can also be about classes, and even about classes containing propositions. For every class of propositions, it then appears that there will have to be a proposition stating that all propositions in that class are true. Let us say that these latter propositions assert the "logical product" of their respective classes.

Intuitively, it seems plausible to imagine that some of these propositions will themselves be in the class whose logical product they assert. Others will not be. For example, the proposition asserting that all propositions in the universal class of propositions are true will itself be a proposition. As a result, it will be in the class whose logical product it asserts. In contrast, the proposition stating that all propositions in the null class of propositions are true will not itself be in the null class of propositions. As a result, it will not be in the class whose logical product it asserts.

Now, consider the class, c_1, which consists of all propositions that state the logical product of some class c_2 in which they are not contained. This class, c_1, will itself be a class of propositions, and so there will exist a proposition — call it P — that states its logical product. A contradiction now arises, for if P is a member of c_1, then it follows that it must not be in c_1; and if it is not in c_1, then it follows that it must be in c_1. As Russell saw it, only by regimenting propositions into a hierarchy of types would this contradiction be avoided.

Both versions of Russell's theory of types eventually came under attack. Some critics claimed the theories were too weak, since they failed to resolve all of the known paradoxes. Other critics claimed they were too strong, since they disallowed many mathematical definitions that, although seemingly consistent, appeared to violate the VCP. Russell's response to the second of these objections was to introduce, within the ramified theory, his so-called axiom of reducibility, an axiom that claimed that any higher-order property or proposition can be reduced to an equivalent first-order one. Although the axiom successfully lessened the VCP's scope of application, many claimed that it was simply too *ad hoc* to be justified philosophically and so, even according to Russell's own list of desiderata, it failed to provide a satisfactory resolution to the paradoxes.

LOGICISM

The discovery of Russell's paradox led Russell to conclude that his work in the *Principles of Mathematics* was at best incomplete and at worst incoherent. Together with Whitehead, he therefore set out to revisit the subject and to provide a more secure foundation for mathematics. The result was *Principia Mathematica*.[10]

Fundamental to *Principia* was the goal of reducing mathematics to logic. In essence, this goal consisted of two main theses. The first was that all mathematical truths should be capable of being translated into logical truths or, in other words, that it should be provable that the vocabulary of mathematics is a proper subset of that of logic. The second was that all mathematical proofs should be capable of being recast as logical proofs or, in other words, that it should be provable that the theorems of mathematics are a proper subset of those of logic.

Russell's basic strategy for defending logicism thus came in two parts. First, there was the idea that numbers should be able to be identified with classes of classes and that classes, in turn, should be able to be identified with propositional functions. Second, there was the idea that number-theoretic statements should be able to be rewritten using the purely logical notions of quantification and identity and then proved as theorems of logic. Thus the number one would be identified with the class of all unit classes, the number two with the class of all two-membered classes, and so on. At the same time, statements such as "There are two books" would be recast as statements such as "There is a book, x, and there is a book, y, and x is not identical to y." It would follow that number-theoretic operations could be explained in terms of logical operations such as conjunction, disjunction and negation and in terms of set-theoretic operations such as intersection, union, and difference.

Like Frege, who had made the same discovery independently sixteen years earlier, Russell observed that counting requires more than a simple one-to-one correspondence between objects and numerals. Since one deck of playing cards can be counted as one deck, four suits, 52 cards, or some other large number of molecules, things can be counted only relative to a given property or predicate (i.e., *qua* decks, suits, cards, etc.). Hence there is the importance of becoming clear on the idea of a propositional function. To say Socrates is an instance of the property of being a man is to say that Socrates is a value of x for which "x is a man" is true. Thus numbers, Russell concludes, will not be properties of things, but of propositional functions. Since, under the right circumstances, classes can be identified with

[10]On the issue of co-authorship, Russell states the following: "The problems with which we had to contend were of two sorts: philosophical and mathematical. Broadly speaking, Whitehead left the philosophical problems to me. As for the mathematical problems, Whitehead invented most of the notation, except in so far as it was taken over from Peano; I did most of the work concerned with series and Whitehead did most of the rest. But this only applies to first drafts. Every part was done three times over. When one of us had produced a first draft, he would send it to the other, who would usually modify it considerably. After which, the one who had made the first draft would put it into final form. There is hardly a line in all the three volumes which is not a joint product." [Russell, 1959, p. 74].

propositional functions, it follows that numbers can be thought of as classes of classes.

According to this view, the number one is a property of any propositional function such that there is an x which makes the function true and, for any y, if y makes the function true, then y turns out to be numerically identical to x. If we call such functions unitary functions, then the number one is the property that all unitary functions share. Similarly, zero is the property of being a null-function (of being a function which is false for all values of x); two is the property of all couples (of being a class for which there are members x and y, such that x is distinct from y and, if z is a member of the class, then z is identical to x or to y); three is the property of all triples, and so on.

More generally, numbers can be said to be classes of classes having the property of similarity. Two classes may be said to be similar if and only if there is a one-to-one, onto relation between their members. For example, in a monogamous country, the number of married men is identical to the number of married women (i.e., there is a one-to-one, onto relation — what Russell calls a "one-one relation" — between these two groups), so the two classes are similar. The number of any class is then defined as the class of all classes similar to it.

These definitions turned out to have several advantages. First they worked mathematically. In other words, they appeared to satisfy all the formal requirements for any theory purporting to give a secure foundation for mathematics. Second, they avoided the philosophical problem of having to postulate zero as a property of nothing, something that had mitigated against the introduction of zero as a number centuries earlier. Third, they worked equally well for both finite and infinite numbers.

As a result, in *Principia Mathematica* Whitehead and Russell were able to provide (in volume 2) many detailed derivations of major theorems in finite and transfinite cardinal arithmetic and in finite and transfinite relation arithmetic. They then went on (in volumes 2 and 3) to develop a general theory of series (including compact, rational and continuous series) and to prove many standard results regarding the limits of functions, the theory of ordinals, the theory of vectors, and elementary measure theory. The derivations, which over three volumes comprised some 1900 pages, are by themselves an immense intellectual achievement; and having achieved so many, and so many elegant and striking, derivations, it is no wonder that Russell and Whitehead became convinced of the truth of logicism. The Introduction (in volume 1, drafted initially by Russell) remains one of the most insightful and informative expositions of logic ever written. A fourth volume (on geometry) was planned but never completed.

Among the most interesting developments due to Russell were his formalization of the intuitive idea of series (in which he defines a series as a set of terms that has an order imposed upon it by a relation that is asymmetrical, transitive, and connected), his novel but quite workable theory of the reals (in which he defines the reals as classes of rational numbers consisting of everything above zero up to a certain point; for example, on this account the square root of two becomes the

class of all rational numbers whose square is less than two), and his extension of mathematical induction into a quite general theory of transfinite induction. Of even greater interest are Russell's comprehensive theory of relations and his amazingly sophisticated theory of relation arithmetic.

Because of his realism, Russell understood relations as something more than merely sets of ordered couples. For Russell, just as a property common to all members of a set gives unity to that set, a relation is what gives unity to sets of ordered couples. It is knowledge of this unity (this property or relation) that allows us to have the knowledge of sets (and of sets of sets, etc.) whose members (whether finite or infinite) we are unable to enumerate.

With this insight in mind, Russell went on to introduce each of the following (now standard) concepts in the theory of relations: (i) the class of terms having relation R to a given fixed term; (ii) the class of terms to which a given fixed term has relation R; (iii) the domain of a relation R, consisting of the class of terms that have R to something or other; (iv) the converse domain of a relation R, consisting of the class of terms to which something or other has R; (v) the field of R, consisting of the union of the domain and converse domain of R; (vi) the converse of a relation R, which is the relation that holds between terms y and x whenever R holds between x and y; (vii) the relative product of two relations, R and S, which holds between x and z whenever there is an intermediate term, y, such that x has R to y and y has S to z; and (viii) plurals, consisting of the class of all terms, if given a separate class c, which have relation R to some member of c.

For example, if R is the relation of being a parent, then the class of terms having R to a given fixed term, a, will be the parents of a; the class of terms to which a given fixed term, b, has relation R will be the children of b; the domain of a relation R will be the class of all people who have children; the converse domain of R will be the class of all people who have parents; the field of R will be the class of all people who either are or have parents; the converse of R will be the relation of being a child (i.e. being a son or daughter); the relative product of the two relations parent and parent will be the relation of being a grandparent; and the relation of being a parent of an Olympian will be an example of a plural.

Using these concepts, Russell went on (in the second half of volume 2) to generalize the idea of an ordinal number. The result was the theory of Russell's so-called "relation numbers." Relation arithmetic then turned out to include many of the familiar formal laws that hold true of ordinary ordinal arithmetic. Beginning with a definition of similarity between classes as the existence of a one-one relation between the classes, Russell defined ordinal similarity (a relation holding between two relations, R and S) as the existence of a correlation between the field of R and the field of S such that, whenever two terms have the relation R, then their correlates also have the relation S, and vice-versa. The relation number of a relation R can then be defined as the class of those relations that are ordinally similar to R. (This is analogous to cardinal arithmetic, but with ordinal similarity being substituted for class similarity and relations being substituted for classes.) Analogous

definitions of the basic arithmetical relations can then be given, such that the new generalized operations of addition and multiplication obey several familiar laws (such as associativity, commutivity and distributivity), although sometimes only in limited cases.

If a well-ordered series is then defined as a series in which any sub-class that has members has a first term, ordinal numbers turn out to be a sub-class of Russell's relation numbers, namely those that apply to well-ordered series. Then, just as Cantor was able to generalize the idea of an ordinal number from the finite to the infinite case, Russell is also able to generalize his idea of a relation number from the finite to the infinite. It was also using relation numbers that Russell was able to provide the first precise definition of the idea of structure, again by generalizing the idea of similarity between classes to that of ordinal similarity with regard to relations.

MATHEMATICAL CERTAINTY

At the same time that he and Whitehead were completing the technical aspects of *Principia Mathematica*, Russell was also focusing his attention on the apparent certainty of mathematical knowledge. Russell saw two tasks as being of fundamental importance for any mathematical epistemology. The first was the task of explaining in what sense "a comparatively obscure and difficult proposition [i.e. a logical axiom] may be said to be a premise for a comparatively obvious proposition [e.g. a truth in elementary arithmetic]."[11] The second was the task of explaining how such comparatively obscure premises are ever discovered and then justified.

In response to the first of these two tasks, Russell distinguishes between two quite different types of premise. The first type is what he calls an "empirical premise," a premise "from which we are actually led to believe the proposition in question."[12] The second is what Russell calls a "logical premise," a premise that is logically simpler than its consequences and that can be used to obtain the proposition in question. (In other words, it is a proposition with fewer logical constituents but powerful enough to generate the kinds of results necessary for a given theory.) Most often in mathematics, Russell then tells us, the empirical and logical premises coincide. It is in exactly these cases that a mathematical proof is of direct epistemological value. However, as Russell points out, this is not always the case. It is simply not true that a logically simpler idea or proposition is always more readily accepted than a more complicated one. Just as is the case with our intuitions about the physical world, it is more commonly the mid-range concepts (concepts which are neither extremely fundamental nor extremely complex) that are usually understood most readily. In some cases, despite their logical simplic-

[11] [Russell, 1907, p. 272].

[12] [Russell, 1907, p. 272]. Russell's use of the phrase "empirical premise" can be misleading. Not all of his "empirical premises" need to be observational, although many are. Nor need they be directly about the empirical world. Rather, what Russell means by an "empirical premise" is simply a premise that has epistemic value.

ity, such premises will have less epistemic simplicity (and less certainty) than the conclusion that follows from them. Hence, there exists the possibility of a "comparatively obscure and difficult proposition" acting as a (logical) premise for a "comparatively obvious proposition."

Russell goes on to note that in these cases it is not the purpose of a mathematical proof so much to prove the conclusion as it is to prove that the conclusion follows from those premises. What such proofs show is that from a particular set of logically simple (but sometimes epistemologically complex) premises, other (sometimes epistemologically simple) conclusions follow deductively. This is important since such proofs help resolve the second of Russell's two tasks, that of explaining how it is that such comparatively obscure and difficult propositions can themselves ever be discovered and justified.

Russell's explanation is that in cases where previously accepted conclusions can be shown to follow from a particular logical premise (or set of premises) via a valid deduction, such a deduction tends to help justify, not the previously accepted conclusion, but rather the original premise (or set of premises) in question. This is a result of what Russell calls the "regressive method."[13] Russell contends that because of this "regressive" aspect of mathematics, the methodology of mathematics is closely related to that of the ordinary sciences of observation. In Russell's words,

> We tend to believe the premises because we can see that their consequences are true, instead of believing the consequences because we know the premises to be true. But the inferring of premises from consequences is the essence of induction; thus the method in investigating the principles of mathematics is really an inductive method, and is substantially the same as the method of discovering general laws in any other science. [Russell, 1907, 1973, pp. 273f]

Science begins with the ordinary facts of observation, of which we are all quite certain. It then attempts to answer two resulting questions: First, what follows from these facts? Second, from what do these facts themselves follow? Answers to the second of these questions determine the general laws of the science, propositions which are logically simpler than observation statements but which are often epistemologically more difficult to justify. When the initial facts are conjoined with these general laws, answers to the first question yield further observation statements, and it is in this way that science gains its predictive power.

According to Russell, mathematics is no exception to this general account. Epistemologically simple propositions such as elementary propositions of arithmetic are originally justified via inference from concrete, often physical, cases.[14] These observations form the basic facts within mathematics of which we are most certain. Statements describing these facts in turn follow from the logically simpler general laws which become as certain as our original empirical premises only if one of the

[13] In a number of respects, Russell's regressive method is similar to Peirce's abduction. In contemporary terminology, it might also be called "inference to the best explanation."

[14] [Russell, 1907, p. 272].

following two cases obtains: either it must be shown that no significantly different hypotheses or general laws could lead to the same empirical premises, or (what is often the case in logic and mathematics) the general laws, once discovered, turn out to be just as obvious as the original empirical premises. As an example of the latter, Russell cites the law of contradiction. This law, Russell feels, "must have been originally discovered by generalizing from instances, though, once discovered, it was found to be quite as indubitable as the instances. Thus it is both an empirical and a logical premise."[15] In the cases where the general laws cannot be shown to be the only ones possible, and are not themselves evident to the same extent as are empirical premises, they must remain merely probable, but often probable to a very high degree. Thus, it is by answering the question, "From what do empirical premises follow?" that Russell's second task (the task of explaining how even epistemologically complex logical premises are discovered and justified) is resolved.

With the above observations in mind, the following summary of Russell's overall theory of mathematical knowledge can be given. According to Russell, mathematical knowledge begins in the first instance from particular observations, e.g., the observation that two objects together with two distinct objects are four objects. These observations form our first epistemologically relevant premises. From these empirical premises we obtain generalizations, e.g., that $2 + 2 = 4$. Such generalizations in turn are often recognized to be "sufficiently obvious to be themselves taken as empirical premises"[16] and so to have additional epistemic value.

In addition to these initial empirical premises of which we are quite certain, there exist two other classes of mathematical knowledge. The first consists of the mathematical knowledge that follows from empirical premises (or from empirical premises together with other known premises) by means of deductive proof. The second consists of that "regressively" justified mathematical knowledge (which includes the general laws of logic and mathematics) from which the original empirical premises can be shown to follow. The first of these two types of mathematical knowledge is reasonably straightforward in terms of its justification. As Russell points out, the expanded body of empirical premises "... when discovered, [is] pretty certain to lead to a number of new results which could not otherwise have been known: in the sciences, this is so obvious that it needs no illustration, and in mathematics it is no less true."[17]

In contrast, the second of these two additional classes of knowledge requires a somewhat more sophisticated account for its justification. Here the general laws of mathematics are discovered "regressively" when even the mathematician inquires after the fewest and logically simplest premises from which all known empirical premises can themselves be deduced. Because of the inductive method that underlies it, such knowledge is sometimes less certain than the (deductively) proven parts of mathematics. As with fundamental laws in the physical sciences,

[15] [Russell, 1907, p. 274].
[16] [Russell, 1907, p. 275].
[17] [Russell, 1907, pp. 282f].

those general laws of logic that do not appear to be as obvious as the original empirical premises will be justified only to the extent that they can be shown to be the most plausible source from which those original premises may be deduced. The result is that even the most fundamental of logical laws may remain merely probable. Russell himself is explicit on this inductivist point:

> In induction, if p is our logical premise and q our empirical premise, we know that p implies q, and in a text-book we are apt to begin with p and deduce q. But p is only believed on account of q. Thus we require a greater or less probability that q implies p, or, what comes to the same thing, that not-p implies not-q. If we can prove that not-p implies not-q, i.e. that p is the only hypothesis consistent with the facts, that settles the question. But usually what we do is to test as many alternative hypotheses as we can think of. If they all fail, that makes it probable, more or less, that any hypothesis other than p will fail. But in this we are simply betting on our inventiveness: we think it unlikely that we should not have thought of a better hypothesis if there were one. [Russell, 1907, pp. 272f]

This inductivist element of Russell's mathematical epistemology differs markedly from the often-repeated account that identifies Russell as a paradigm advocate of the view that mathematical knowledge gains its certainty directly as a result of its deductive foundations. This often-repeated view, when it notices Russell's inductivist tendencies at all, usually puts off these tendencies as the occasional lapses of an otherwise stanch defender of what might be called "epistemic logicism."

Such an interpretation is understandable. After all, other logicists have conspicuously advocated positions which are decidedly inconsistent with any inductive or "regressive" element whatsoever. Frege, for example, espouses the view that axioms are truths, as are theorems, but that "they are truths for which no proof can be given in our system, and for which no proof is needed. It follows that there are no false axioms, and that we cannot accept a thought as an axiom if we are in doubt about its truth ... ."[18] In contrast, for Russell it is logic that ultimately gains its justification as a result of its mathematical consequences, not vice-versa, and it is the distinction between logical and epistemological order within mathematics that Russell emphasizes time and time again throughout his career. For example, in the first volume of *Principia* we read the following:

> But in fact self-evidence is never more than a part of the reason for accepting an axiom, and is never indispensable. The reason for accepting an axiom, as for accepting any other proposition, is always largely inductive, namely that many propositions which are nearly indubitable can be deduced from it, and that no equally plausible way is known by which these propositions could be true if the axiom were false, and nothing which is probably false can be deduced from it. If the axiom is

[18] [Frege, 1914, p. 205].

> apparently self-evident, that only means, practically, that it is nearly indubitable; for things have been thought to be self-evident and have yet turned out to be false. And if the axiom itself is nearly indubitable, that merely adds to the inductive evidence derived from the fact that its consequences are nearly indubitable: it does not provide new evidence of a radically different kind. Infallibility is never attainable, and therefore some element of doubt should always attach to every axiom and to all its consequences. In formal logic, the element of doubt is less than in most sciences, but it is not absent, as appears from the fact that the paradoxes followed from premisses which were not previously known to require limitations. [Whitehead and Russell, 1910, p. 62] (second edn, p. 59).

Later, in the Introduction to the second edition in 1927, Russell's comments are to much the same effect when he introduces talk of the "purely pragmatic justification" of the axiom of reducibility.[19]

Précis of this same position are given in several of Russell's other publications. In his *Introduction to Mathematical Philosophy* Russell observes that the propositions of simple arithmetic are more obvious than those of logic and that "The most obvious and easy things in mathematics are not those that come logically at the beginning; they are things that, from the point of view of logical deduction, come somewhere in the middle."[20] Later, in his 1924 essay "Logical Atomism", Russell again explains his position but in greater detail. His comments are worth quoting in their entirety because of their clarity:

> When pure mathematics is organized as a deductive system ... it becomes obvious that, if we are to believe in the truth of pure mathematics, it cannot be solely because we believe in the truth of the set of premises. Some of the premises are much less obvious than some of their consequences, and are believed chiefly because of their consequences. This will be found to be always the case when a science is arranged as a deductive system. It is not the logically simplest propositions of the system that are the most obvious, or that provide the chief part of our reasons for believing in the system. With the empirical sciences this is evident. Electro-dynamics, for example, can be concentrated into Maxwell's equations, but these equations are believed because of the observed truth of certain of their logical consequences. Exactly the same thing happens in the pure realm of logic; the logically first principles of logic — at least some of them — are to be

[19] [Whitehead and Russell, 1910, p. xiv]. Despite such comments, Russell apparently never gave up the hope of deducing such axioms from other, more self-evident logical truths. For example, see the Introduction to the second edition of *Principia*, vol. 1, p. xiv. Once it is admitted, as Russell does, that these axioms are in part empirical, such hope seems inexplicably misguided since, if this is so, it follows immediately that one would not expect them to be derivable from purely logical premises, whether self-evident or not.

[20] [Russell, 1919, p. 2].

> believed, not on their own account, but on account of their consequences. The epistemological question: 'Why should I believe this set of propositions?' is quite different from the logical question: 'What is the smallest and logically simplest group of propositions from which this set of propositions can be deduced?' Our reasons for believing logic and pure mathematics are, in part, only inductive and probable, in spite of the fact that, in their logical order, the propositions of logic and pure mathematics follow from the premises of logic by pure deduction. I think this point important, since errors are liable to arise from assimilating the logical to the epistemological order, and also, conversely, from assimilating the epistemological to the logical order. [Russell, 1924, pp. 361f]. Reprinted in [Russell, 1956, pp. 325f]

Just what are the lessons that should be learned from such comments? The major lesson is that Russell's regressive method, emphasizing as it does the distinction between logical and epistemological order, shows how closely Russell's mathematical epistemology was integrated within his general theory of knowledge. Wood, for example, emphasizes exactly this close relationship. After describing the importance of inductive reasoning for Russell, Wood concludes that the regressive method is characteristic, not only of Russell's mathematical epistemology, but of his epistemology in general.[21] Russell himself bolsters this contention when he states that throughout his academic life, epistemological concerns have been the "one constant preoccupation"[22] and when he notes that Wood's account "admirably clarifies various things that might otherwise cause misunderstanding."[23]

At the same time, upon careful reading it becomes clear that Russell's regressive method is not in any way inconsistent with other comments of Russell's, such as the claim that "Mathematics is a deductive science."[24] Read in context, it becomes clear in such cases that Russell is referring only to the logical order of the discipline. Following the above quotation, for example, Russell goes on to explain that "No appeal to common sense, or 'intuition', or anything except strict deductive logic, ought to be needed in mathematics *after* the premises have been laid down",[25] leaving unanswered within the then current context the question of just how the original axioms are to be obtained.

The second major lesson that should be learned from the above observations is just how important Russell felt the analogy between epistemological concerns in mathematics and in the sciences to be. Given the number of times that Russell emphasizes this analogy, together with the fact that the stated purpose of these comments was "to emphasize the close analogy between the methods of pure mathematics and the methods of the sciences of observation",[26] Russell's intention

[21] [Wood, 1959, pp. 265–266].
[22] [Russell, 1959, p. 11].
[23] [Russell, 1959, p. 3].
[24] [Russell, 1919, p. 144].
[25] [Russell, 1919, p. 145] (emphasis added).
[26] [Russell, 1907, p. 272].

should be clear: only by emphasizing this analogy can a complete and accurate picture of the acquisition and nature of mathematical knowledge be obtained.

LOGIC AND ANALYTIC PHILOSOPHY

In much the same way that Russell used his new logic in an attempt to clarify issues in mathematics, he also used it in an attempt to clarify issues in philosophy. As one of the founders of analytic philosophy, Russell made significant contributions to a wide variety of areas, including metaphysics, epistemology, ethics and political theory, as well as to the history of the discipline. Underlying these various projects were not only Russell's use of logical analysis, but also his long-standing aim of discovering whether, and to what extent, any kind of knowledge is possible. "Philosophy arises from an unusually obstinate attempt to arrive at real knowledge," he writes in 1927. "What passes for knowledge in ordinary life suffers from three defects: it is cocksure, vague, and self-contradictory."[27] As a result, Russell took the question of how much can be known, and with what degree of certainty or doubtfulness, to be the most essentially philosophical of all questions.

More than this, Russell's various contributions to both mathematics and philosophy were also unified by his views concerning both the centrality of scientific knowledge and the importance of an underlying scientific methodology common to philosophy and science. In the case of philosophy, this methodology expressed itself through Russell's use of logical analysis. In fact, Russell often claimed that he had more confidence in his methodology than in any particular philosophical conclusion.

Russell's methodology consisted of the making and testing of hypotheses through the weighing of evidence (hence Russell's comment that he wished to emphasize the "scientific method" in philosophy), together with a rigorous analysis of problematic propositions using the machinery of first-order logic. It was Russell's belief that by using the new logic of his day, philosophers would be able to exhibit the underlying "logical form" of natural language statements. A statement's logical form, in turn, would help philosophers resolve problems of reference associated with the ambiguity and vagueness of natural language. The result was Russell's doctrine of logical atomism (the doctrine that the world consists of a plurality of independently existing particulars exhibiting qualities and standing in relations, with all complex facts being constructed out of these logically simple building blocks) and the complete integration of logic with metaphysics.

Just as we distinguish three separate sense of "is" (the *is* of predication, the *is* of identity, and the *is* of existence) and exhibit these three senses by using three separate logical notations ($Px, x = y$, and $\exists x$ respectively), we will also discover other ontologically significant distinctions by being aware of a sentence's correct logical form. On Russell's view, the subject matter of philosophy is then distinguished from that of the sciences only by the generality and the *a prioricity*

[27] [Russell, 1927, p. 1].

of philosophical statements, not by the underlying methodology of the discipline. In philosophy, as in mathematics, Russell believed that it was by applying logical machinery and insights that advances would be made.

Russell's most famous example of his "analytic" method concerns denoting phrases such as descriptions and proper names. In his *Principles of Mathematics*, Russell held the view that every denoting phrase (for example, "Scott," "blue," "the number two," "the golden mountain") denoted, or referred to, an existing entity. By the time his landmark article, "On Denoting," appeared two years later, in 1905, Russell had modified this extreme realism and had instead become convinced that denoting phrases need not possess a theoretical unity.

While logically proper names (words such as "this" or "that" which refer to sensations of which an agent is immediately aware) necessarily have referents associated with them, descriptive phrases (such as "the largest number less than pi") should be viewed simply as a collection of quantifiers (such as "all" and "some") and propositional functions (such as "x is a number"). As such, they need not be understood as referring terms but, rather, as "incomplete symbols." In other words, such phrases may be thought of as symbols that take on meaning within appropriate contexts, but that are meaningless in isolation.

Thus, in the sentence "The present King of France is bald," the definite description "The present King of France" plays a role quite different from that of a proper name such as "Scott" in the sentence "Scott is bald." For example, letting K abbreviate the predicate "is a present King of France," B abbreviate the predicate "is bald," and s abbreviate the name "Scott," Russell represents the former as $\exists x[(Kx \ \& \ \forall y(Ky \supset y = x)) \ \& \ Bx]$ and the latter simply as Bs. This distinction between distinct logical forms allows Russell to explain away three important puzzles.

The first of these puzzles relates to true negative existential claims, such as the claim "The golden mountain does not exist." Here, once again, by treating definite descriptions as having a logical form distinct from that of proper names, Russell is able to give an account of how a speaker may be committed to the truth of a negative existential without also being committed to the belief that the subject term has reference. That is, the claim that Scott does not exist is false since the sentence "$\sim \exists x(x = s)$" is self-contradictory. (After all, there must be at least one thing that is identical to s since it is a logical truth that s is identical to itself!) In contrast, the claim that a golden mountain does not exist may be true since, assuming that G abbreviates the predicate "is golden" and Mabbreviates the predicate "is a mountain," there is nothing contradictory about "$\sim \exists x(Gx \ \& \ Mx)$."

The second puzzle concerns the operation of the Law of Excluded Middle and how this law relates to denoting terms. According to one reading of the Law of Excluded Middle, it must be the case that either "The present King of France is bald" is true or "The present King of France is not bald" is true. But if so, both sentences appear to entail the existence of a present King of France, clearly an undesirable result whenever no such individual exists. Russell's analysis shows

how this conclusion can be avoided. By appealing to the above analysis, it follows that there is a way to deny the sentence "The present King of France is bald" without being committed to the existence of a present King of France, namely by accepting that "It is not the case that there exists a present King of France who is bald" is true.

The third puzzle concerns the Law of Identity as it operates in (so-called) opaque contexts. Even though "Scott is the author of Waverley" is true, it does not follow that the two referring terms "Scott" and "the author of Waverley" are interchangeable in every situation. Thus although "George IV wanted to know whether Scott was the author of Waverley" is true, "George IV wanted to know whether Scott was Scott" is, presumably, false. Russell's distinction between the logical forms associated with the use of proper names and definite descriptions shows why this is so. To see this we once again let s abbreviate the name "Scott." We also let w abbreviate "Waverley" and A abbreviate the two-place predicate "is the author of." It then follows that the sentence "$s = s$" is not at all equivalent to the sentence "$\exists x[Axw \ \& \ \forall y(Ayw \supset y = x) \ \& \ x = s]$."

Russell's emphasis upon logical analysis had immediate consequences for his metaphysics. In response to the traditional problem of the external world which, it is claimed, arises since the external world can be known only by inference, Russell developed his famous 1910 distinction between "knowledge by acquaintance and knowledge by description." He then went on, in his 1918 lectures on logical atomism, to argue that the world itself consists of a complex of logical atoms (such as "little patches of colour") and their properties. Together they form the atomic facts which, in turn, are combined to form logically complex objects. What we normally take to be inferred entities (for example, enduring physical objects) are then understood to be "logical constructions" formed from the immediately given entities of sensation, viz., "sensibilia." For Russell, it is only these latter entities that are known non-inferentially and with certainty.

Put in other words, it is the philosopher's job is to discover a logically ideal language that will exhibit the true nature and structure of the world in such a way that the speaker will not be misled by the casual surface structure of natural language. Just as atomic facts (the association of universals with an appropriate number of individuals) may be combined into molecular facts in the world itself, such a language would allow for the description of such combinations using logical connectives such as "and" and "or." In addition to atomic and molecular facts, Russell also held that general facts (facts about "all" of something) were needed to complete the picture of the world. Famously, he vacillated on whether negative facts were also required.

Thus, unlike Leibniz, Russell believed there was more to scientific, mathematical and philosophical knowledge than 'getting the concepts right.' Even so, just as Leibniz had wanted to develop a logically ideal language (or *characteristica universalis*), together with an instrument of universal deductive reasoning (his *calculus ratiocinator*), Russell believed that modern logic would help us represent the world as it actually is and reason about it without fear of error. As Russell

himself translates one of Leibniz's most famous quotations,

> We should be able to reason in metaphysics and morals in much the same way as in geometry and analysis ... If controversies were to arise, there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and to say to each other (with a friend as witness, if they liked): Let us calculate. [Russell, 1900, pp. 169-170]

Logical analysis may not have the air of empirical discovery about it, but it is just as crucial to the scientific enterprise as observation and experiment.

CONCLUSION

By contemporary standards, Russell's logic is in some respects quite primitive. The notation is clumsy. Many of the crucial technical ideas were borrowed from Peano, and some of the most important philosophical ideas from Leibniz and Poincaré. Over the years, the axioms that Russell and Whitehead introduced have been simplified in significant ways. The theory of types is unbearably complicated and scholars still argue about whether Russell's use of propositional functions was sophisticated or confused. There appears to be no significant appreciation of the explosion of meta-theoretical results that would soon follow Russell's groundbreaking insights.

But regardless of such shortcomings, a century of hindsight has proved the lasting value of Russell's work. The now archaic notation (introduced largely by Whitehead) was in its day a tremendous improvement upon that of Frege and Peano, and it is unlikely that without it the field of modern logic would have gained the prominence it has today. New proposals for the simplification of axioms and the theory of types came as much from Russell and his students as from anyone else. The most fertile seeds for modern work in meta-theory and model theory have been found more often in the questions Russell posed about his own work than anywhere else.

Even more importantly, it was Russell's logic that prompted important new research to be done in fields as diverse as philosophy, mathematics, economics, decision theory, computability theory and computer science. In December 1955, when Herbert Simon and Allen Newell wanted to test their first "thinking machine," it was to Russell's logic that they immediately turned. After developing the first theorem prover designed to replicate many of the theorems of *Principia*, the two men wrote to Russell to tell him of their accomplishment. "I am delighted to know that *Principia Mathematica* can now be done by machinery," Russell wrote back. I wish Whitehead and I had known of this possibility before we wasted 10 years doing it by hand."[28] As Isaac Newton reminds us, intellectual

[28] [Lewis, 2001].

and scientific progress only occurs if we are privileged enough to be able to stand on the shoulders of giants.

BIBLIOGRAPHY

[Blackwell and Ruja, 1994] K. Blackwell and H. Ruja. *A Bibliography of Bertrand Russell*, 3 vols, London: Routledge, 1994.

[Broad, 1973] C. D. Broad. Bertrand Russell, as Philosopher, *Bulletin of the London Mathematical Society*, 5, 328-341, 1973; repr. in A.D. Irvine, *Bertrand Russell: Critical Assessments*, vol. 1, New York and London: Routledge, 1999, 1-15.

[Carnap, 1931] R. Carnap. The Logicist Foundations of Mathematics, *Erkenntnis*, 2, 91-105, 1931; repr. in Benacerraf, Paul, and Hilary Putnam (eds), *Philosophy of Mathematics*, 2nd ed., Cambridge: Cambridge University Press, 1983, 41-52; in Klemke, E.D. (ed.), *Essays on Bertrand Russell*, Urbana: University of Illinois Press, 1970, 341-354; in Pears, David F. (ed.), *Bertrand Russell: A Collection of Critical Essays*, Garden City, New York: Anchor Books, 1972, 175-191; and ; in A.D. Irvine, *Bertrand Russell: Critical Assessments*, vol. 2, New York and London: Routledge, 1999, 135-146.

[Church, 1974] A. Church. Russellian Simple Type Theory, *Proceedings and Addresses of the American Philosophical Association*, 47, 21-33, 1974.

[Church, 1976] A. Church. Comparison of Russell's Resolution of the Semantical Antinomies with That of Tarski, *Journal of Symbolic Logic*, 41, 747-760, 1976; repr. in A.D. Irvine, *Bertrand Russell: Critical Assessments*, vol. 2, New York and London: Routledge, 1999, 96-112.

[Clark, 1975] R. W. Clark. *The Life of Bertrand Russell*, London: J. Cape, 1975.

[Clark, 1981] R. W. Clark. *Bertrand Russell and His World*, London: Thames and Hudson, 1981.

[Copi, 1971] I. Copi. *The Theory of Logical Types*, London: Routledge and Kegan Paul, 1981.

[Eames, 1969] E. R. Eames. *Bertrand Russell's Theory of Knowledge*, London: George Allen and Unwin, 1969.

[Frege, 1893, 1903] G. Frege. Grundgesetze der Arithmetik, 2 vols, Jena: Pohle, 1893, 1903; abridged and translated by Montgomery Furth as *The Basic Laws of Arithmetic*, Berkeley: University of California Press, 1964.

[Frege, 1902] G. Frege. Letter to Russell, 1902. In van Heijenoort, Jean, *From Frege to Gödel*, Cambridge, Mass.: Harvard University Press, 1967, 126-128.

[Frege, 1903] G. Frege. The Russell Paradox, 1903. In *The Basic Laws of Arithmetic*, Berkeley: University of California Press, 1964, 127-130; abridged and repr. in A.D. Irvine, *Bertrand Russell: Critical Assessments*, vol. 2, New York and London: Routledge, 1999, 1-3.

[Frege, 1914] G. Frege. Logic in Mathematics, 1914. Translated by Peter Long and Roger White in Frege, *Posthumous Writings*, Chicago: University of Chicago Press, 1979, 203-250.

[Gandy, 1973] R. O. Gandy. Bertrand Russell, as Mathematician, *Bulletin of the London Mathematical Society*, 5, 342-348, 1973; repr. in A.D. Irvine, *Bertrand Russell: Critical Assessments*, vol. 1, New York and London: Routledge, 1999, 16-23.

[Gödel, 1944] K. Gödel. Russell's Mathematical Logic. In Schilpp, Paul Arthur (ed.), *The Philosophy of Bertrand Russell*, 1944. 3rd ed., New York: Tudor, 1951, 123-153; repr. in Benacerraf, Paul, and Hilary Putnam (eds), *Philosophy of Mathematics*, 2nd ed., Cambridge: Cambridge University Press, 1983, 447-469; in Pears, David F. (ed.) (1972) *Bertrand Russell: A Collection of Critical Essays*, Garden City, New York: Anchor Books, 192-226; and; in A.D. Irvine, *Bertrand Russell: Critical Assessments*, vol. 2, New York and London: Routledge, 1999, 113-134.

[Grattan-Guinness, 1977] I. Grattan-Guinness. *Dear Russell, Dear Jourdain: A Commentary on Russell's Logic, Based on His Correspondence with Philip Jourdain*, New York: Columbia University Press, 1977.

[Griffin, 1991] N. Griffin. *Russell's Idealist Apprenticeship*, Oxford: Clarendon, 1991.

[Hager, 1994] P. J. Hager. *Continuity and Change in the Development of Russell's Philosophy*, Dordrecht: Nijhoff, 1994.

[Hylton, 1990a] P. W. Hylton. Logic in Russell's Logicism. In Bell, David, and Neil Cooper (eds), *The Analytic Tradition: Philosophical Quarterly Monographs*, Vol. 1, Cambridge: Blackwell, 137-172, 1990.

[Hylton, 1990b] P. W. Hylton. *Russell, Idealism, and the Emergence of Analytic Philosophy*, Oxford: Clarendon, 1990.

[Irvine, 1989] A. D. Irvine. Epistemic Logicism and Russell's Regressive Method, *Philosophical Studies*, 55, 303-327, 1989.

[Irvine, 1999] A. D. Irvine, ed. *Bertrand Russell: Critical Assessments*, 4 vols, London: Routledge, 1999.

[Irvine, 2004] A. D. Irvine. Russell on Method. In Godehard Link (ed.), One Hundred Years of Russell's Paradox, Berlin, New York: Walter de Gruyter, 481-500, 2004.

[Irvine and Wedeking, 1993] A. D. Irvine and G. A. Wedeking, eds. *Russell and Analytic Philosophy*, Toronto: University of Toronto Press, 1993.

[Jager, 1972] R. Jager. *The Development of Bertrand Russell's Philosophy*, London: George Allen and Unwin, 1972.

[Kaplan, 1970] D. Kaplan. What is Russell's Theory of Descriptions? In Yourgrau, Wolfgang, and Allen D. Breck, (eds), *Physics, Logic, and History*, New York: Plenum, 277-288, 1970; repr. in Pears, David F. (ed.), B*ertrand Russell: A Collection of Critical Essays,* Garden City, New York: Anchor Books, 1972, 227-244.

[Klemke, 1970] E. D. Klemke. *Essays on Bertrand Russell*, Urbana: University of Illinois Press, 1970.

[Landini, 1998] G. Landini. *Russell's Hidden Substitutional Theory*, New York and Oxford: Oxford University Press, 1998.

[Lewis, 2001] P. Lewis. Herbert A. Simon, Nobel Winner for Economics, Dies at 84, *New York Times*, 10 February 2001.

[Linsky, 1999] B. Linsky. *Russell's Metaphysical Logic*, Stanford: CSLI Publications, 1999.

[Lycan, 1981] W. Lycan. Logical Atomism and Ontological Atoms, *Synthese*, 46, 207-229, 1981.

[Monk, 1996] R. Monk. *Bertrand Russell: The Spirit of Solitude*, London: Jonathan Cape, 1996.

[Monk, 2000] R. Monk. *Bertrand Russell: The Ghost of Madness*, London: Jonathan Cape, 2000.

[Monk and Palmer, 1996] R. Monk and A. Palmer, eds. *Bertrand Russell and the Origins of Analytic Philosophy*, Bristol: Thoemmes Press, 1996.

[Nakhnikian, 1974] G. Nakhnikian, ed. *Bertrand Russell's Philosophy*, London: Duckworth, 1974.

[Patterson, 1993] W. Patterson. *Bertrand Russell's Philosophy of Logical Atomism*, New York: Lang, 1993.

[Pears, 1967] D. F. Pears. *Bertrand Russell and the British Tradition in Philosophy*, London: Collins, 1967.

[Pears, 1972] D. F. Pears, ed. *Bertrand Russell: A Collection of Critical Essays*, New York: Doubleday, 1972.

[Putnam, 1967] H. Putnam. The Thesis that Mathematics is Logic. In Schoenman, Ralph (ed.), *Bertrand Russell: Philosopher of the Century*, London: Allen and Unwin, 273-303, 1967; repr. in Putnam, Hilary, Mathematics, Matter and Method, Cambridge: Cambridge University Press, 1975, 12-42.

[Quine, 1938] W. V. O. Quine. On the Theory of Types, *Journal of Symbolic Logic*, 3, 125-139, 1938; repr. in A.D. Irvine, *Bertrand Russell: Critical Assessments*, vol. 2, New York and London: Routledge, 1999, 78-95.

[Quine, 1966] W. V. O. Quine. *Selected Logic Papers*, New York: Random House, 1966.

[Ramsey, 1926] F. P. Ramsey. Mathematical Logic, *Mathematical Gazette*, 13, 185-194, 1926; repr. in Ramsey, Frank Plumpton, *The Foundations of Mathematics*, London: Kegan Paul, Trench, Trubner, 1931, 62-81; in Ramsey, Frank Plumpton, *Foundations*, London: Routledge and Kegan Paul, 1978, 213-232; in Ramsey, Frank Plumpton, *Philosophical Papers*, Cambridge: Cambridge University Press, 1990, 225-244; and in A.D. Irvine, *Bertrand Russell: Critical Assessments*, vol. 2, New York and London: Routledge, 1999, 54-66.

[Ramsey, 1960] F. P. Ramsey. *The Foundations of Mathematics*, Paterson, NJ: Littlefield, Adams and Co, 1960.

[Roberts, 1979] F. W. Roberts, ed. *Bertrand Russell Memorial Volume*, London: Allen and Unwin, 1979.

[Rodrigues-Consuegra, 1991] F. A. Rodriguez-Consuegra. *The Mathematical Philosophy of Bertrand Russell: Origins and Development*, Basel: Birkhauser Verlag, 1991.
[Russell, 1897] B. Russell. *An Essay on the Foundations of Geometry*, Cambridge: At the University Press, 1897.
[Russell, 1900] B. Russell. *A Critical Exposition of the Philosophy of Leibniz*, Cambridge: At the University Press, 1900.
[Russell, 1901] B. Russell. Recent Work on the Principles of Mathematics, *International Monthly,* 4, 83-101, 1901; repr. as "Mathematics and the Metaphysicians" in Russell, Bertrand, *Mysticism and Logic*, London: Longmans Green, 1918, 74-96.
[Russell, 1902] B. Russell. Letter to Frege, 1902. In van Heijenoort, Jean, *From Frege to Gödel*, Cambridge, Mass.: Harvard University Press, 1967, 124-125.
[Russell, 1903] B. Russell. *The Principles of Mathematics*, Cambridge: At the University Press, 1903.
[Russell, 1905] B. Russell. On Denoting, *Mind*, 14, 479-493, 1905; repr. in Russell, Bertrand, *Essays in Analysis*, London: Allen and Unwin, 1973, 103-119.
[Russell, 1907] B. Russell. The Regressive Method of Discovering the Premises of Mathematics, 1907. In *Essays in Analysis*, London: George Allen and Unwin, 1973, 272-283.
[Russell, 1908] B. Russell. Mathematical Logic as Based on the Theory of Types, *American Journal of Mathematics*, 30, 222-262, 1908; repr. in Russell, Bertrand, *Logic and Knowledge,* London: Allen and Unwin, 1956, 59-102, and in van Heijenoort, Jean, *From Frege to Gödel*, Cambridge, Mass.: Harvard University Press, 1967, 152-182.
[Russell, 1910] B. Russell. Knowledge by Acquaintance and Knowledge by Description, *Proceedings of the Aristotelian Society*, 11, 108-128, 1910; repr. in Russell, Bertrand, *Mysticism and Logic*, London: Allen and Unwin, 1963, 152-167.
[Russell, 1910] B. Russell. *Philosophical Essays*, London: Longmans, Green, 1910.
[Russell, 1912a] B. Russell. On the Relations of Universals and Particulars, *Proceedings of the Aristotelian Society*, 12, 1-24, 1912; repr. in Russell, Bertrand, *Logic and Knowledge*, London: Allen and Unwin, 1956, 105-124.
[Russell, 1912b] B. Russell. *The Problems of Philosophy*, London: Williams and Norgate; New York: Henry Holt and Company, 1912.
[Russell, 1914] B. Russell. *Our Knowledge of the External World*, Chicago and London: The Open Court Publishing Company, 1914.
[Russell, 1918] B. Russell. *Mysticism and Logic and Other Essays*, London and New York: Longmans, Green, 1918; repr. as *A Free Man's Worship and Other Essays*, London: Unwin Paperbacks, 1976.
[Russell, 1918/1919] B. Russell. The Philosophy of Logical Atomism, *Monist*, 28, 495-527; 29, 32-63, 190-222, 345-380, 1918, 1919; repr. in Russell, Bertrand, *Logic and Knowledge*, London: Allen and Unwin, 1956, 177-281.
[Russell, 1919] B. Russell. *Introduction to Mathematical Philosophy*, London: George Allen and Unwin; New York: The Macmillan Company, 1919.
[Russell, 1924] B. Russell. Logical Atomism. In Muirhead, J.H., *Contemporary British Philosophers*, London: Allen and Unwin, 1924, 356-383; repr. in Russell, Bertrand, *Logic and Knowledge,* London: Allen and Unwin, 1956, 323-343.
[Russell, 1927] B. Russell. *An Outline of Philosophy*, London: George Allen and Unwin, 1927; repr. as Philosophy, New York: W.W. Norton, 1927.
[Russell, 1928] B. Russell. *Sceptical Essays*, London: George Allen and Unwin; New York: W.W. Norton, 1928.
[Russell, 1931] B. Russell. *The Scientific Outlook*, London: George Allen and Unwin; New York: W.W. Norton, 1931.
[Russell, 1940] B. Russell. *An Inquiry into Meaning and Truth*, London: George Allen and Unwin; New York: W.W. Norton, 1940.
[Russell, 1948] B. Russell. *Human Knowledge: Its Scope and Limits*, London: George Allen and Unwin; New York: Simon and Schuster, 1948.
[Russell, 1949] B. Russell. *The Philosophy of Logical Atomism*, Minneapolis, Minnesota: Department of Philosophy, University of Minnesota, 1949; repr. as *Russell's Logical Atomism,* Oxford: Fontana/Collins, 1972.
[Russell, 1956] B. Russell. *Logic and Knowledge: Essays, 1901-1950*, London: George Allen and Unwin; New York: The Macmillan Company, 1956.
[Russell, 1959] B. Russell. *My Philosophical Development*, New York: Simon and Schuster, 1959.

[Russell, 1967–69] B. Russell. *The Autobiography of Bertrand Russell*, 3 vols, Boston and Toronto: Little Brown and Company (Vols 1 and 2), 1967, 1968, New York: Simon and Schuster (Vol. 3), 1969.
[Russell, 1973] B. Russell. *Essays in Analysis*, London: George Allen and Unwin, 1973.
[Russell, 1984] B. Russell. *Theory of Knowledge: The 1913 Manuscript* (The Collected Papers of Bertrand Russell, Vol. 7), London, Boston, Sydney: George Allen and Unwin, 1984.
[Russell, 1986] B. Russell. *The Philosophy of Logical Atomism and Other Essays* (The Collected Papers of Bertrand Russell, Vol. 8), 1914-19, London: George Allen and Unwin, 1986.
[Russell, 1992] B. Russell. *Logical and Philosophical Papers, 1909-13* (The Collected Papers of Bertrand Russell, Vol. 6), London and New York: Routledge, 1992.
[Russell, 1994a] B. Russell. *Foundations of Logic, 1903-05* (The Collected Papers of Bertrand Russell, Vol. 4), London and New York: Routledge, 1994.
[Russell, 1994b] B. Russell. *Toward the Principles of Mathematics* (The Collected Papers of Bertrand Russell, Vol. 3), London and New York: Routledge, 1994.
[Aavage and Anderson, 1989] C. W. Savage and C. A. Anderson, eds. *Rereading Russell: Essays on Bertrand Russell's Metaphysics and Epistemology*, Minneapolis: University of Minnesota Press, 1989.
[Schlipp, 1944] P. A. Schilpp, ed. *The Philosophy of Bertrand Russell*, Chicago: Northwestern University, 1944; 3rd ed., New York: Harper and Row, 1963.
[Schoenman, 1967] R. Schoenman, ed. *Bertrand Russell: Philosopher of the Century*, London: Allen and Unwin, 1967.
[Slater, 1994] J. G. Slater. *Bertrand Russell*, Bristol: Thoemmes, 1994.
[Strawson, 1950] P. F. Strawson. On Referring, *Mind*, 59, 320-344, 1950; repr. in Flew, Anthony (ed.), *Essays in Conceptual Analysis*, London: Macmillan, 1960, 21-52, and in Klemke, E.D. (ed.), *Essays on Bertrand Russell*, Urbana: University of Illinois Press, 1970, 147-172.
[Urquhart, 1988] A. Urquhart. Russell's Zig-Zag Path to the Ramified Theory of Types, *Russell*, 8, 82-91, 1988.
[Weitz, 1944] M. Weitz. Analysis and the Unity of Russell's Philosophy. In Schilpp, Paul Arthur (ed.), *The Philosophy of Bertrand Russell*, 1944; 3rd ed., New York: Tudor, 1951, 55-121.
[Whitehead and Russell, 1910–1913] A. N. Whitehead and B. Russell. *Principia Mathematica*, 3 vols, Cambridge: Cambridge University Press, 1910, 1912, 1913; second edition, 1925 (Vol. 1), 1927 (Vols 2, 3).
[Wittgenstein, 1921] L. Wittgenstein. *Logisch-philosophische Abhandlung*, 1921. Trans. as *Tractatus Logico-Philosophicus*, London: Kegan Paul, Trench, Trubner, 1922.
[Wittgenstein, 1956] L. Wittgenstein. *Remarks on the Foundations of Mathematics*, Oxford: Blackwell, 1956.
[Wood, 1957] A. Wood. *Bertrand Russell: The Passionate Sceptic*, London: Allen and Unwin, 1957.
[Wood, 1959] A. Wood. Russell's Philosophy: A Study of its Development. In Russell, Bertrand, *My Philosophical Development*, New York: Simon and Schuster, 255-277, 1959.

LOGIC FOR MEINONGIAN OBJECT THEORY SEMANTICS

Dale Jacquette

1 MEINONG ON INTENDED OBJECTS

Alexius Meinong (1853–1920) was a talented musician, historian, psychologist, and philosopher — but not much of a logician. His student Ernst Mally was a highly competent logician, and served as a liaison for Meinong between current developments in formal symbolic logic and the school of intentional experimental psychology and semantic theory that Meinong founded in Graz, Austria, at the turn of the previous century. We cannot, as a result, speak of Meinong's logic, if we wish to be historically accurate, but at most only of various versions of Meinongian logic, inspired by Meinong's object theory (*Gegenstandstheorie*), and intended to formalize the principles by which Meinong sought to implement his teacher Franz Brentano's thesis of the distinctive intentionality of thought.[1]

What Meinong wanted to achieve was a theory of intended objects, most of which, barring only a few interesting exceptions, were thinkable, but in general independent of actual psychological acts. Meinong agrees with Brentano that every thought is directed toward an intended object, but he denies Brentano's early immanent intentionality thesis, formulated in the latter's monumental 1874 *Psychologie vom empirischen Standpunkt*, according to which the intended objects of thought are literally contained within the acts of thought by which they are intended. Meinong, like others of Brentano's students, preferred to distance himself from this aspect of his teacher's philosophy, which was interpreted as amounting to a kind of philosophically objectionable psychologism. Instead, Meinong, in collaboration with Alois Höfler in Höfler's 1890 *Logik*, advanced a distinction that was to be developed at greater length in Kazimierz Twardowski's 1894 *Zur Lehre vom Inhalt und Gegenstand der Vorstellung*, between the act, content, and intended object of every psychological occurrence. Where Brentano in retrospect seems to have conflated content and object, Höfler, Meinong, Twardowski, and others in this early phenomenological tradition, argued instead that there is a sharp distinction between content and object, and that whereas the content of thought is

[1] A detailed account of the philosophical background to the formalization of Meinong's object theory would require many times the number of pages for this effort to sketch one logical theory of Meinongian semantics. The references below offer an indication of the vast literature on the subject.

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

something the thought immanently contains, the intended object of a thought is not generally a part of the thought by which it is intended, but something that stands outside of thought.[2]

In this way, Meinong and company proposed to establish a mind-independent order of intended objects to which thought might be directed, but that were not themselves in any sense the creatures of thought. The objects (*Gegenstände*) of Meinong's object theory were thus meant to facilitate a rapprochement between a reconstructed Brentanian thesis of the intentionality of mind with commonsense reference to the objects intended in psychological acts of perception and memory, where we ordinarily think of ourselves as directed in thought toward things that exist outside of our own thoughts, and believe ourselves to be capable of thinking of the same external objects as other persons. Thus, when I think about the Blue Mosque in Istanbul, I suppose myself to be directed in thought to an entity that exists outside of the contents of my own mind in real space and time, in a certain part of far southeastern Europe, that would exist even if I were never to think of it, and that stands as the very same intended object of thoughts of other thinkers who are also not merely intending the contents of their own distinct mental episodes when like me they also think about the Blue Mosque.

By virtue of adopting a qualified form of Brentano's intentionality thesis, Meinong nevertheless opened up a vista on a wonderful domain of intended objects, only some of which like the Blue Mosque happen to exist. If every thought intends an object, and if intended objects are not psychologistically to be confused with thought contents, but are in some sense independent of any individual acts of thinking, then Meinong, true to these assumptions, must somehow accommodate within his theory the intended objects of such thoughts as 'Pegasus is a winged horse', 'The golden mountain is golden and a mountain', 'Priestley believed in Phlogiston', and even, 'The round square is round and square'.

Each of these thoughts on Meinong's amended form of Brentano's intentionality thesis ostensibly intends a nonexistent object. The examples are different in interesting ways. The first, concerning Pegasus, seems to be about a mythological flying equine, something that, as far as we know, happens contingently never to have existed. The second example, about the golden mountain, apparently expresses a kind of tautology or analytic truth. It is comparable in form to saying of a contingently existent object, such as the tallest mountain, 'The tallest mountain is tall and a mountain'. If this thought and its expression in language are true, even logically or analytically true, then why should another thought or sentence of identical logical form not equally be true, merely because it intends an object that happens not to exist? What can pure logic pretend to know about the kinds of mountains that do or do not exist in the world? This is nevertheless the con-

[2]See Dale Jacquette, "The Origins of *Gegenstandstheorie*: Immanent and Transcendent Intentional Objects in Brentano, Twardowski, and Meinong", *Brentano Studien*, 3, 1990–1991, 277–302. Jacquette, "Twardowski, Brentano's Dilemma, and the Content-Object Distinction", in *Actions, Products, and Things: Brentano and Polish Philosophy*, Mind and Phenomenology series, edited by Arkadiusz Chrudzimski and Dariusz Lukasiewicz, pp. 9–33 (Frankfurt: Ontos Verlag, 2006.

clusion that extensionalist logic and semantics since the time of Gottlob Frege and Bertrand Russell after 1905, through the influential work of W.V.O. Quine and others in mainstream analytic philosophy, have vigorously maintained with a desperate conviction bordering almost on religious fervor.[3]

The third example is interesting in another way, because it combines reference to an existent object, the 17th century chemist Joseph Priestley, with what has been judged by science to be the nonexistent object, the substance of Phlogiston, invoked before the discovery of oxygen in order to explain combustion and oxidation. Here Meinong and common sense in one of its moods will declare that the proposition is true, that Priestley did indeed (mistakenly) believe in the existence of Phlogiston, something that only later turns out not to have actually existed. If it is true of Preistly that he believed in the existence of Phlogiston, however, then in order for this true thought and its expression in language to be true, we must presumably be able to refer to and truly predicate properties of a nonexistent intended object, namely Phlogiston. This nonexistent stuff, wrongly supposed to be given off when something burns or rusts, has not only these properties, for that is how Phlogiston is defined and understood as distinct from other existent and nonexistent chemical substances, but has the more remarkable property of having been believed by the existent chemist Priestley to exist. Priestley was wrong, we now know, his thought was false, but it was a thought *about* Phlogiston, and that, indeed, is the essential part of what makes his thought false.

Finally, the fourth example is interesting for Meinong's purposes because it is identical in logical form to the second example, but it concerns the apparently analytic attribution of properties to a putatively intended object that not only

[3]Bertrand Russell is responsible for much of the misinformation concerning Meinong's object theory semantics that has prevailed in the analytic philosophical community beginning with Russell's justly celebrated 1905 *Mind* essay, "On Denoting". See Russell, *Introduction to Mathematical Philosophy* (New York: Simon and Schuster, 1971), p. 169. Russell's *Introduction* was first published in 1919, after his conversion to radical extensionalism. But Russell's interpretation of Meinong's object theory as committed to the being-predication thesis begins with his earliest critical commentaries. Russell, "Meinong's Theory of Complexes and Assumptions" (*Mind*, 13, 1904), rpt., *Essays in Analysis*, edited by Douglas Lackey (New York: George Braziller, 1973), p. 36: "The process suggested by Meinong's argument is...exceedingly and curiously complicated. First we think of a golden mountain, then we perceive that we are thinking of it; thence, we infer that there is a presentation of a golden mountain, and thence finally that the golden mountain subsists or has Being." Also p. 59: "The immanent object does not exist, according to Meinong, and is therefore no part of the mental state whose object it is; for this mental state exists. Yet, although not part of any mental state, it is supposed to be in some sense psychical. But it cannot be in any way bound up with any particular mental state of which it is the object; for other states, at other times and in other people, may have precisely the same object, since an object or a proposition can be presented or believed more than once. I confess these facts seem to me to show, without more ado, that objects and propositions must always have being..." See Jacquette, "Confessions of a Meinongian Logician", *Grazer Philosophische Studien*, 58–59, 2000, pp. 151–180, for some debunking of Meinong-bashing. The conference on 'Russell versus Meinong: 100 Years After "On Denoting"', organized by myself and Nicholas Griffin at McMaster University in Hamilton, Ontario, Canada, 14-18 May 2005, may have also helped dispel the myths that have surrounded Meinong's philosophy seen through Russell's eyes throughout most of the history of modern symbolic logic.

contingently happens not to exist, like Pegasus, Phlogiston, or the golden mountain, but an object that by definition cannot possibly exist. The round square we may suppose is round and square at once, and it is for this very reason, by virtue of possessing these properties, that there cannot be a round square. Still, if someone were to ask what the fourth example listed above was about, the correct answer would seem to be that it involved a thought about the round square, that *that* was the object intended, even though no such thing exists or can exist.

2 MEINONG'S PHENOMENOLOGICAL TAXONOMY OF OBJECTS

Trained by Brentano to apply empirical methods to philosophical investigations, Meinong pursues a policy similar to that of a naturalist studying mushrooms or butterflies. He collects all the relevant examples and tries to organize them into a sensible set of categories according to their distinctive features.

Thus, Meinong distinguishes between intended objects that have being and those that do not have being. He observes a medieval Scholastic distinction that he probably inherited from Brentano's detailed study of Thomas Aquinas between existent and subsistent entities as objects with being. Existent objects on this distinction are objects with being that occupy real space and time; they are physically existent. Subsistent objects in contrast are abstract objects with being, in which category Meinong includes numbers, but interestingly not geometrical figures like the triangle, and properties, qualities and relations, and propositions as the abstract meanings of concrete sentences. These are also assimilated to the category of states of affairs under Meinong's concept of *Objektive* (objectives), themselves a subcategory of intended objects. Most contemporary logicians who have been interested in Meinong's object theory have not been motivated to capture the existence–subsistence distinction in their formalizations of Meinongian logic, nor shall we in what follows below.

Among the objects lacking being or possessing nonbeing, Meinong as a semantic taxonomist identifies as further types those that are incomplete (*unvollständige Gegenstände*) and those that are impossible (*unmögliche Gegenstände*). Incomplete objects include all those intended objects that are contingently nonexistent, such as Pegasus, the golden mountain, Phlogiston, Santa Claus, Sherlock Holmes, and others, that are nevertheless beingless intended objects of thought. These objects are incomplete beingless objects according to Meinong because they are lacking at least some property and its complement. Thus, unlike any truly existent object, Pegasus is neither blue-eyed nor non-blue-eyed, neither an Arabian nor a non-Arabian, and so on. Similarly for all other contingently beingless objects; they are all lacking some, indeed, countlessly many, properties in a way that is never true of actually existent or subsistent entities as intended objects of thought. It is worth remarking that Meinong's reason for regarding generalized geometrical figures like the triangle as lacking being because he accepts an argument of the eighteenth century British empiricist George Berkeley in his *Three Dialogues Between Hylas and Philonous*, to the effect that we can have no idea of 'the triangle'

if by this we are supposed to mean an object that is neither scalene nor isosceles, right-angled nor non-right-angled, neither large nor small, with line segments that are neither red nor white nor any other particular color, and so on. Whether this is a conclusive basis for excluding the triangle as an abstract mathematical entity from the order of intended objects with being rather depends on what kinds of properties we expect such an object needs in order to qualify as complete rather than incomplete.

Impossible objects are also typically but not necessarily incomplete. The two subcategories of beingless objects accordingly are not exclusionary but permit a substantial overlap. The round square is a good example with which to begin. For, like the triangle, it is neither red nor non-red, neither large nor small, and so on. It is only simultaneously both round and square. This implies in effect that the round square is both round and non-round, square and non-square, which no object with being can possibly have. It is not the kind of thing like a golden mountain or a flying horse that might eventually turn up somewhere in the universe, or whose existence does not offend the assumptions of ontology, but something that by definition could not exist because its properties cannot jointly be realized in any existent object. We might nevertheless suppose that there could be an impossible object that was also complete, if, for example, such an object were defined as possessing every property and the complement of at least one of its properties, so that it would be just like an existent object except that, if, for example, it were red, it would also be non-red; or if it were round, it would also be non-round.

Meinong, again in keeping with sound empiricist methodology, also proposes to extract laws or principles governing the intended objects in the domain of objects he posits as a consequence of his revisionary Brentanian intentionality thesis. He maintains in the simplest most basic part of the theory with which we will be concerned in this exposition that every assumption *intends* an object; that the population of a semantic domain by intended objects is *indifferent* to their ontic status as having or failing to have being; that the so-being (*Sosein*), being-thus-and-so, or possession of constitutive properties of an intended object is *independent* of its ontic status. What is characteristic of an intensional Meinongian semantics on this account and in contrast with extensionalist semantics is that it permits reference and true predication of constitutive properties to beingless as well as existent spatiotemporal and subsistent abstract intended objects. Semantically, a Meinongian object theory is extraordinarily democratic; all objects are on an equal footing at least insofar as they can stand as the objects of reference and true predications of properties. The golden mountain and round square do not exist, and the round square in particular cannot possibly exist. We can nevertheless think of them, and, like existent objects, they are among the intended objects of thought and language to which we can refer. Despite their ontic status as beingless, they truly possess the constitutive properties predicated of them in assumptions and related sorts of mental acts.[4]

[4]For formal definitions of incomplete and impossible objects, see Jacquette, *Meinongian Logic: The Semantics of Existence and Nonexistence* (Berlin: Walter de Gruyter, 1996), p. 117, and

3 MEINONGIAN LOGIC FOR MEINONG'S OBJECT THEORY SEMANTICS

A Meinongian logic of existent and nonexistent objects provides a formal theory of reference and true predication for ordinary discourse, the semantics of ontological commitment, and logic of fiction. The Meinongian logic proposed in what follows offers a rigorous Meinongian object theory semantics with nonstandard propositional and predicate inference machinery. All such efforts are necessarily reconstructive, for the reasons we have already mentioned. Meinong was no logician, but he had interesting insights into the nature of intensional semantics that were fueled by his commitment to some version of Brentano's thesis of the intentionality of thought, and derivatively to the intentionality of language as the expression of thought.

Every logician who has tried to work out a Meinongian logic, and at this point in time there have been about a dozen, have recognized the need to pick and choose from among Meinong's ideas, to idealize his suggestions and synthesize them into a form that more or less conforms to the kinds of things Meinong explicitly said about the meanings of terms, concerning the reference and true predication of properties to beingless objects. An historically faithful contemporary symbolic logic that formalizes Meinong's *Gegenstandstheorie* is unavoidably conjectural and a matter of approximation. That is the beauty and the temptation of trying to mathematicize Meinongian semantics as a reconstructive task that must choose among various elements of Meinong's theory as he developed it over time, changing his opinions along the way. Meinong did not write with the logician in mind, but with his finger on the beating pulse of intentionalist phenomenological psychology, and with a finely-tuned sensitivity to the nuances of thought and language to which a semantics in his opinion must always pay attention. It is for this reason that Meinong's writings provide a wealth of useful raw material for the construction of a new kind of logic, a logic that Meinong himself was never himself in a position to articulate.

The system presented below is distinguished from previous formalizations of object theory by virtue of its formal criteria for nuclear (constitutive) and extranuclear (nonconstitutive) properties, three-valued propositional semantics for predications of nuclear properties to incomplete nonexistent objects for which the objects ostensibly are undetermined, nonstandard set theory semantics with unrestricted comprehension for object theory predicate semantics (licensed by existence restrictions on abstraction equivalence), demonstrations of internal determinacy, consistency, and Henkin completeness, nonstandard deduction theorem, and consistency considerations in light of free assumption and unrestricted comprehension. It is a version of a Meinongian logic I have presented on several previous occasions, with new refinements and applications. Much of what appears in the formalization will seem familiar to classical logicians, and that is just the point. A Meinongian

for theorems relating these and other properties of basic Meinongian categories of objects, pp. 122-140.

logic need not be so entirely different from the kinds of extensionalist logics already available in many variations, including three-valued logics with some of the innovations appropriated here for intensionalist purposes. By exhibiting the possibility of providing a consistency, completeness, and compactness proof for Meinongian logics with restricted lambda-abstraction and other precautions, the present system seeks to demonstrate that Meinongian logics are formally as rigorous as their extensionalist existence-presupposing counterparts, and as entitled, therefore, to take their place alongside other nonstandard logics in the field.

4 ELEMENTS OF A RECONSTRUCTED MEINONGIAN OBJECT THEORY

Meinongian object theories, in the revisionary sense intended here, can be constructed around the following core of basic principles.

1. Any thought or corresponding expression can be assumed.
   (Principle of unrestricted free assumption or *Annahmen* thesis)

2. Every assumption is directed toward an intentional object.
   (Intentionality thesis)

3. Every intentional object has a nature, character, *Sosein*, 'how-it-is', 'so-being', or 'being thus-and-so', regardless of its ontological status.
   (Independence or independence of *Sosein* from *Sein* thesis)

4. Being or nonbeing is not part of the *Sosein* of any intentional object, nor of the object considered in itself.
   (Indifference thesis or doctrine of the *Aussersein* of the homeless pure object)

5. There are two modes of being or *Sein* for intentional objects:

   (a) Spatio-temporal existence

   (b) Platonic subsistence

   (*Existenz*/*Bestand* thesis)

6. There are some intentional objects which do not have *Sein*, but neither exist nor subsist (objects of which it is true to say that there are no such objects).

As indicated, most object theories that are Meinongian in all other respects do not recognize or make use of the distinction between existence and subsistence. For present purposes it is appropriate to relax principle (5) in order to permit theories of this kind also to qualify as Meinongian. Thesis (6) is similarly expendable when derived from (1)–(4). Meinong's exposition of the theory includes additional postulates, but the present amended list of assumptions is sufficient for

a minimal Meinongian object theory. Alternative versions can be formulated by supplementing the core with further assumptions consistent with it.[5]

5 SYNTAX, FORMATION AND INFERENCE PRINCIPLES

The formalization provides an exact interpretation of object theory logic and semantics. The system developed is ontologically neutral in areas where traditional logic is intolerably partisan. It is more encompassing and flexible than alternative standard and nonstandard logics of comparable power. The logic is paradox-free and untyped. It preserves intact almost all the notation and inferences of standard systems, replacing the usual propositional semantics with three truth values and a Meinongian domain of existent and nonexistent objects.

5.1 Nuclear and Extranuclear Properties

The distinction between nuclear and extranuclear properties is represented in the notation as a distinction between 'P' and '$P!$' predicates. Reference to nuclear or extranuclear properties indifferently is indicated by placing the exclamation in parentheses, '$P(!)$'. Quantification over nuclear or extranuclear properties is indicated by expressions of the form:

$$\ldots(\forall x)\ldots(\forall P^n(!))\ldots(\ldots P^n(!)x_1,\ldots,x_n\ldots)\ldots$$

The predications are higher-order in appearance only. The logic is type-unordered, and propositions of the kind are instances of bounded quantification, convenient abbreviations for the more explicitly type-unordered expression $\ldots(\forall x)\ldots(\forall y(!))\ldots(\ldots y(!)x\ldots)\ldots$. The leveling of types is appropriate in Meinongian logic because properties like anything that can be thought of are also intentional objects.

Nuclear P properties and extranuclear $P!$ properties are distinguished by the following criteria:

(C1) $\quad -(\forall x_1)\ldots(\forall x_n)(\forall P^n)(-P^n x_1,\ldots,x_n \equiv \overline{P^n}x_1,\ldots,x_n)$

(C2) $\quad (\forall x_1)\ldots(\forall x_n)(\forall P^n!)(-P^n!x_1,\ldots,x_n \equiv \overline{P^n}!x_1,\ldots,x_n)$

These can also be called the negation/counterpart-complementation criteria for the nuclear-extranuclear property distinction. It suffices to rely on (C1) or (C2)

[5]See Alexius Meinong, *Alexius Meinong Gesamtausgabe*, Vols. I–VIII, edited by Rudolf Haller and Rudolf Kindinger in collaboration with R. M. Chisholm (Graz: Akademische Druck– u. Verlagsanstalt, 1969–1978). Also J.N. Findlay, *Meinong's Theory of Objects and Values*, *edited with an Introduction by Dale Jacquette*, from the second edition, Oxford University Press, 1963, Aldershot, Ashgate Publishing (Gregg Revivals) (all references to this edition). Karel Lambert, *Meinong and the Principle of Independence: Its Place in Meinong's Theory of Objects and its Significance in Contemporary Philosophical Logic* (Cambridge: Cambridge University Press, 1983). Alternative formalizations of object theory are advanced by Terence Parsons, *Nonexistent Objects* (New Haven: Yale University Press, 1980), and Richard Routley, *Exploring Meinong's Jungle and Beyond*, interim edition (Canberra: Australian National University, 1981).

exclusively, categorizing extranuclear properties as properties that do not satisfy (C1), or nuclear properties as properties that do not satisfy (C2).

The criteria mark an important distinction between nuclear and extranuclear properties. Nuclear property and property complement pairs can fail to hold of incomplete or indeterminate nonexistent objects. George Berkeley's golden mountain (to belabor a familiar example) is neither taller than nor not taller than René Daumal's Mt. Analogue. Criterion (C1) reflects the fact that the property of being taller than Mt. Analogue is nuclear rather than extranuclear by denying the logical equivalence of sentence negation and nuclear predicate complementation counterpart expressions. To say that the round square is nonsquare (because it is round) is *not* to say that it is not the case that the round square is square (since by the independence thesis it is both round *and* square).[6]

Extranuclear properties on the contrary do not permit this latitude even when predicated of nonexistent incomplete indeterminate or impossible objects. An object, whether determinate or indeterminate, is either existent or not existent, possible or impossible, self-identical or not self-identical, with no middle ground or room for extranuclear incompleteness or indeterminacy. This means, as (C2) in effect requires, that extranuclear predicate complementation and sentence negation counterparts are logically equivalent and extensionally intersubstitutable. For this reason, the extranuclear subtheory of Meinongian logic, unlike the object theory of nuclear predications, is classically bivalent. Nuclear constitutive properties belong to the distinguishing or uniquely individuating *Sosein* of an intentional object, to the absolute exclusion of extranuclear properties.

Criteria (C1) and (C2) by themselves do not determine which properties are nuclear and which extranuclear. The criteria state distinguishing features of the two kinds of properties, and in application require additional information about whether or not the complement of a particular property predicated of an object is logically equivalent to its sentence negation counterpart. An independent informal criterion for the distinction can be provided to determine intuitively which properties satisfy (C1) rather than (C2), and which (C2) rather than (C1). A property is extranuclear on this conception if and only if it is definable in terms of logical operators and uninterpreted predicate symbols alone. It is nuclear if and only if it is not extranuclear, or if it is not definable in this way, but requires for its definition the interpretation of at least some predicate symbols.

5.2 Syntax

Object theory O contains the following items of primitive syntax.

[6] Meinong introduces the distinction between *konstitutorische* and *ausserkonstitutorische Bestimmungen* (constitutive and extraconstitutive properties) in *Über Möglichkeit und Wahrscheinlichkeit: Beiträge zur Gegenstandstheorie und Erkenntnistheorie* (Leipzig: Verlag von Johann Ambrosius Barth, 1915) (same pagination as Barth), pp. 176-177. Findlay, in *Meinong's Theory of Objects and Values*, p. 176, proposes the English translations 'nuclear' and 'extranuclear'. See also Parsons, "Nuclear and Extranuclear Properties, Meinong, and Leibniz", *Noûs*, 12, 1978, pp. 137-151. Parsons, *Nonexistent Objects*.

1. propositional negation: '$-$'
2. predicate complementation: '$\overline{}$'
3. conditional: '$\supset$'
4. universal quantification: '$\forall$'
5. object variable: 'x'
6. object constant: 'o'
7. predicates: 'P^n' (nuclear); '$P^n!$' (extranuclear)
   ('$P^n(!)$' denotes an n-ary nuclear or extranuclear property indifferently; in practice the superscript is implicit)
8. functor: 'f^n'
9. denumerable variable, constant, predicate, and functor indices:
   '1', '2', '3', ...
10. punctuation, including:

    parentheses: '(', ')'

    square brackets: '[', ']'

    curly brackets: '{', '}'

    comma: ','

    ellipsis: '...'

    blank space: ' '

In addition to the primitive syntax, defined expressions, including connectives, predicates, functors, and operators, are introduced as abbreviations or particular applications of more complex expressions in the primitive syntax. In principle, the primitive syntax can supply all needed variables, constants, predicates, and functors, by subscripting natural number indices to the 'x', 'o', '$P(!)$, and 'f' expressions, as in: 'x_1', 'x_2', 'x_3', ..., 'o_1', 'o_2', 'o_3', ..., '$P(!)_1$', '$P(!)_2$', '$P(!)_3$',..., and 'f_1', 'f_2', 'f_3', ... In practice it is more convenient to use variables like 'x', 'y', 'z', predicates like 'P', '$Q!$', '$R(!)$', or those with mnemonic significance, and function symbols from traditional logical and mathematical notations, such as '$\in$', '$\leq$', '$\neq$'.

Object constants are given special interpretations for convenience and readability. In particular, an object constant $o_i (1 \leq i \leq n)$ may be chosen to represent an arbitrary proposition, though in practice familiar propositional variables 'p', 'q', and 'r' are used. Three object constants are chosen to represent the truth values 'T', 'F', and 'U'. They are understood as the values true, false, undetermined. These conventional applicationss are translatable into the official primitive syntax.

5.3 Formation Principles

The terms of object theory O are recursively inductively defined as follows.

1. The $x_1 \ldots, x_n, \ldots$ are terms.
2. The $o_1, \ldots, o_n, \ldots$ are terms.
3. If f^n is n-ary, then $f^n(o_1, \ldots, o_n)$ is a term.
4. Nothing else is a term (closure).

The well-formed formulas or wffs of object theory O are recursively inductively defined as follows.

1. If $P^n(!)$ is n-ary, then $P^n(!)o_1 \ldots o_n$ is a wff.
2. If $P^n(!)$ is n-ary, then $\overline{P^n}(!)o_1 \ldots o_n$ is a wff.
3. If p is a wff, then $-p$ is a wff.
4. If p and q are wffs, then $p \supset q$ is a wff.
5. If p is a wff, then $(\forall x)p$ is a wff.
6. Nothing else is a wff (closure).

The following connectives and operators are introduced as abbreviations for the primitive connectives and operators to which they are logically equivalent.

E1 $p \,\&\, q$ for $-((-p \supset -q) \supset -q)$

E2 $p \vee q$ for $-(-p \,\&\, -q)$

E3 $p \equiv q$ for $(p \supset q) \,\&\, (q \supset p)$

E4 $(\exists x)p$ for $-(\forall x) - p$

The formation principles can also be expanded to incorporate modality, definite description, and lambda abstraction.

5.4 Inference Principles

A proof in the logic is a consecutively numbered sequence of well-formed formulas. The inferences are justified by analogues of familiar axioms and rules of propositional and predicate logic.

The intent of inference structures is to permit the deduction from true assumptions of true conclusions only, and to prevent the deduction of false or undetermined conclusions. For present purposes, it will suffice to advance a number of logical axioms generally recognized as deductively valid, from which derived inference principles can be obtained. The turnstile '$\vdash$' is used to indicate the valid

deduction of a conclusion q from logical and nonlogical axioms, definitions, and theorems $p_1, \ldots, p_n$ in $p_1, \ldots, p_n \vdash q$. This means that it is deductively valid to infer q from $p_1, \ldots, p_n$. In a proof this is written:

$$\begin{array}{rll} 1. & p_1 & \\ & \vdots & \\ n. & p_n & \\ n+1 & q & (1-n) \end{array}$$

The turnstile ($\vdash$) is a metatheoretical device, a (non-truth-functional) function from propositions to propositions. When the turnstile appears with a proposition to its right but none to its left, it indicates that the proposition is unconditionally true, a theorem or tautology of the logic.

The primitive axioms are presented as inference schemata, uniform substitution instances of which may enter into any proof. Inference is by uniform substitution of propositions for propositional variables, and a rule of detachment, *modus ponendo ponens*.

A1 $p \supset (q \supset p)$

A2 $(p \supset q) \supset ((q \supset r) \supset (p \supset r))$

A3 $(-p \supset p) \supset p$

A4 $(-p \supset -q) \supset (q \supset p)$

A5 $(\forall x_1) \ldots (\forall x_n) P^n(!)x_1 \ldots x_n \vdash P^n(!)x_i/o_i$ (similarly for $\ldots \overline{P^n}(!) \ldots$)
(selectively substituting constants 'o_i' for variables 'x_i')

A6 $P^n(!)x_i/o_i \vdash (\forall x_1) \ldots (\forall x_n) P^n(!)x_1 \ldots x_n$ (similarly for $\ldots \overline{P^n}(!) \ldots$)
(where 'x_i' does not occur free in '$P^n(!)x_i/o_i$', '$P^n(!)x_i/o_i$')

A7 $(\forall x_1) \ldots (\forall x_n)(p \supset q) \supset (p \supset (\forall x_1) \ldots (\forall x_n q)$
(where 'x_i' does not occur free in 'p')

A8 $(\forall x_1) \ldots (\forall x_n)(P^n(!)x_1 \ldots x_n \equiv \overline{\overline{P^n}}(!)x_1 \ldots x_n)$

The logic distinguishes between propositional negation and predicate complementation. This requires a special double complementation principle in (A8). The axiom states that an arbitrary predication of given length involving the complement of the complement of a property is equivalent to a predication of the same length of the property itself. If it is not the case that an object has a nuclear property, it does not follow that the object has the complement of the property. If an object has the complement of a nuclear property, it does not follow that the object does not have the property. The importance of these restrictions in maintaining syntactical consistency are appreciated when the logic of incomplete and impossible objects is presented.

6 OBJECT THEORY SEMANTICS

The intended interpretation of the logic is described by explaining the meaning of each term and well-formed formula. The semantics divides into two parts. If p, q are wffs of O, then truth matrix interpretations are provided for negation in $-p$, and implication in $p \supset q$. If p is of the form $P^n(!)o_1 \ldots o_n$ or $(\forall x)p$, it is interpreted in a semantic model, expressed as an ordered pair of reference class designators, or as an equivalent function on the Cartesian product of objects in designated reference classes.

The most important primitive semantic relation is designation, not of existents only, but of any object or set of objects in the semantic domain of existent and nonexistent Meinongian objects. The dictum attributed to Quine that there can be 'No entity without identity'[7] is expanded to provide the more inclusive principle that there can be no entity *or* nonentity without identity. Nonexistent objects have *Sosein* identity conditions specifiable as identities of unordered sets of nuclear properties. Existent and nonexistent objects alike are distinguishable elements of the Meinongian semantic domain, and can therefore stand as the values of bound variables (which in Meinongian quantificational logic carry no real existential or ontological import). This removes philosophical obstacles to the claim that nonexistent objects are designatable.

The domain or universe of discourse includes a wide variety of objects, existent as well as incomplete and impossible nonexistent objects. Meinong identified the major divisions among kinds of intentional objects, but he did not recognize all interesting subdivisions. These can be mapped as intersections, unions, and exclusions of partitioned reference classes of nonexistent objects.

There are three truth values in the logic's formal semantics. They are required because intentional discourse sometimes attributes properties to incomplete objects in predications that intuitively are neither true nor false. A third truth value is therefore introduced for expressions in the logic undetermined in truth value. The golden mountain is neither taller than nor not taller than Daumal's Mt. Analogue. Macbeth neither spoke nor was unable to speak Italian. Truth values are distinguished as special objects of the semantic domain, and an identity predicate is used to express identity relationships between truth values and truth value functions applied to propositions.

6.1 Three-Valued Propositional Semantics

Let p, q be any well-formed formulas. Three objects in the semantic domain are designated as truth values, symbolized in the usual way as 'T' for 'true', 'F' for 'false', and 'U' for 'undetermined'. A function V is defined for wffs of the

[7] The thesis is ascribed to Quine by Charles Parsons, "Frege's Theory of Number", in Max Black, editor, *Philosophy in America* (Ithaca: Cornell University Press, 1965), p. 182. See W.V.O. Quine, *Word and Object* (Cambridge: The MIT Press, 1960), pp. 200–205; Quine, "Ontological Relativity", in Quine, *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), pp. 26–68, pp. 32–34, 45–64.

logic that gives the truth value of any wff taken as argument. If $V(p) = T$ then $V(-p) = F$; if $V(p) = T$ and $V(q) = F$, then $V(p \supset q) = F$; $V(p) = U$ if and only if $V(-p) = U$. The complete truth functional relations for negation and the conditional are represented in truth tables.

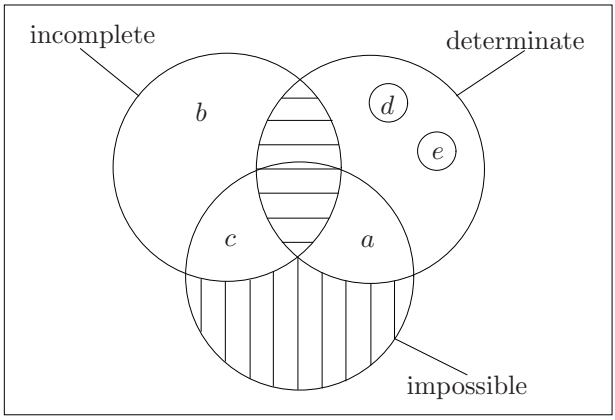
p	$-q$
T	F
F	T
U	U

p	q	$p \supset q$
T	T	T
T	F	F
T	U	U
F	T	T
F	F	T
F	U	T
U	T	T
U	F	U
U	U	T

The undetermined truth value U is preserved through negation and through conjunction with anything that has truth value T. It is natural to suppose that if a wff is undetermined in truth value, then its negation is also undetermined. If it is undetermined that the golden mountain is taller than Mt. Analogue, then it is also undetermined that the golden mountain is not taller than Mt. Analogue. If one conjunct of a conjunction is true and the others undetermined, then the conjunction as a whole cannot unqualifiedly be said to be true or false. The undetermined truth value is not preserved through conjunction with any proposition that has truth value F, since a conjunction is rightly regarded as false if even one of its conjuncts is false. The consequences of these evaluations for the conditional are recorded in the truth tables. The propositional semantics are identical to the trivalent system proposed by Jan Łukasiewicz.[8]

Predicate expressions of the form $\overline{P^n}(!)o_1 \ldots o_n$ and complementary form $P^n(!)o_1 \ldots o_n$ are interpreted by an extension of the truth value function. It is necessary to describe a domain of existent and nonexistent objects. Among nonexistent objects are impossible, incomplete, and impossible incomplete objects, overdetermined objects, and a maximally impossible object. These are represented diagrammatically on the following page.

[8] Jan Łukasiewicz, "Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls", *Comptes rendus des seánces de la Société des sciences et des lettres de Varsovie*, Classe III, 23, 1930, pp. 51–77; translated in Storrs McCall, editor, *Polish Logic: 1920–1939* (Oxford: Oxford University Press, 1967), pp. 40–65. See Nicholas Rescher, *Many–Valued Logic* (New York: McGraw–Hill, 1969), p. 23. The truth tables are similar to those presented by S.C. Kleene, *Introduction to Metamathematics* (Amsterdam: North–Holland Publishing Co., 1952), pp. 334–335. The difference is that for Lukasiewicz, when $V(p) = U$, $V(p \supset p) = T$; whereas for Kleene, when $V(p) = U, V(p \supset p) = U$.



In this hybrid Venn-Euler diagram, intentional objects are found within every sphere, and no objects occur outside the interlocking spheres. Clearly, no objects belong to the intersection of determinate and incomplete objects, nor to the domain of neither incomplete nor determinate impossible objects, and in the diagram this is indicated by hatching. Object a is the maximally impossible object which has every nuclear constitutive property and its complement. Object b is the (incomplete fictional) golden mountain, and c the (impossible) round square. Subsets d and e might distinguish Meinong's ontological categories of existent and subsistent determinate objects, though the distinction need not be observed.

6.2 Meinongian Set Theory

To interpret '$P^n(!)o_1 \dots o_n$' and '$\overline{P^n}(!)o_1 \dots o_n$', a semantic model is provided that can be directly transcribed from the logical syntax. The model consists of three ordered components: a domain D, an interpretation I on the domain, and a truth valuation of propositions under the interpretation V, $\langle D, I, V \rangle$. D is the Meinongian domain of existent and nonexistent intentional objects, $D = \{o_1, o_2, o_3, \dots\}$. Every term designates an existent or nonexistent object, by the *Annahmen* or unrestricted free assumption thesis. Domain D thus contains an intentional object corresponding to every well-defined term, constant, predicate, functor (and definite descriptor and lambda abstract, if these are added). Some of the objects are sets of objects, where every condition on any objects determines a set.

A nonstandard version of Zermelo-Fraenkel set theory drives the semantics. Meinongian set membership, inclusion and exclusion, are nuclear relational properties in atomic predications undetermined in truth value. The philosophical justification for this is that incomplete objects like Pegasus intuitively are undetermined with respect to nuclear properties not included in their *Soseine* or so-beings. The proposition that Pegasus is blue-eyed intuitively is neither true nor false, so that the corresponding membership statement that Pegasus $\in$ {blue-eyed things} must also be undetermined.

The classical set identity axiom states, for sets φ, ψ, membership '$\in$' undefined:

$$(\forall\varphi)(\forall\psi)[\varphi = \psi \equiv (\forall x)(x \in \varphi \equiv x \in \psi)]$$

The principle cannot be introduced without modification in trivalent semantics, since it can happen that it is true or false that an object is a member of set φ but undetermined that the object is a member of set ψ, or the reverse. The truth matrix in that case dictates that it is also undetermined that $\varphi = \psi$, though intuitively identity unlike set membership is a bivalent extranuclear relation.

There are several ways of refining the standard extension axiom for set identity to preserve bivalence of identity for three-valued membership. The most straightforward method is to require that membership of an object in sets φ and ψ always has the same truth value when and only when set identity holds. The amended set identity axiom states:

$$(\forall\varphi)(\forall\psi)(\varphi = \psi \equiv (\forall x)[(x \in \varphi \equiv x \in \psi)\&V(x \in \varphi) = V(x \in \psi)])$$

This assures that identity predications for sets are always true or false and never undetermined, precluding identity when it is true or false that an object is a member of φ, but undetermined that it is a member of ψ, or the reverse.

The principle may appear to be circular because of its explicit reference to the valuations of membership predications in the second conjunct, since the truth values of predicate expressions are about to be interpreted in terms of set theoretical operations on the domain of objects. Vicious circularity is avoided by the consideration that interpretations of predicate expressions in the logic presuppose but do not explicitly involve set identities. There is no more circularity in the nonstandard set identity axiom than in standard applications of set theory in the semantics of predicate logic, where predicate and quantificational expressions are interpreted in set theoretical models, and the axioms of set theory are themselves expressed as predicate or quantificational expressions. The loophole here as in standard semantics lies in a hierarchy of logical object and metalanguages. The semantics are given in a metalanguage to interpret predicate expressions in a subordinate object language; the metalanguage is expressed by means of its own higher-order object language, interpreted in an even higher-order metametalanguage.

The appearance of circularity in the nonstandard set identity axiom is entirely avoided by introducing a new propositional operator '$\Leftrightarrow$' according to the following matrix, collapsing three truth values to two:

Strong Equivalence

p	q	$p \Leftrightarrow q$
T	T	T
T	F	F
T	U	F
F	T	F
F	F	T
F	U	F
U	T	F
U	F	F
U	U	T

This makes it possible to redefine the Meinongian set identity axiom without hint of predicate valuation circularity.

$$(\forall\varphi)(\forall\psi)[\varphi = \psi \equiv (\forall x)(x \in \varphi \Leftrightarrow x \in \psi)]$$

The complete axioms of the Meinongian set theory underlying the predicate semantics can now be given; again for any sets φ, ψ, membership '$\in$' and the null set '$\varnothing$' undefined.

Concept of Set
$(\forall x)[Set(x) \equiv (\exists y)(y \in x \vee x = \varnothing)]$

(Nuclear-Unrestricted) Comprehension
$(\exists\varphi)(\forall P)(\forall x)(x \in \varphi \equiv Px)$

Meinongian Set Identity
$(\forall\varphi)(\forall\psi)(\varphi = \psi \equiv (\forall x)[(x \in \varphi \equiv x \in \psi)\,\&\,V(x \in \varphi) = V(x \in \psi)])$
$(\forall\varphi)(\forall\psi)[\varphi = \psi \equiv (\forall x)(x \in \varphi \Leftrightarrow x \in \psi)]$

Subsets
$(\forall x)(\forall P(!))(\exists\varphi)[x \in \varphi \equiv P(!)x\,\&\,(\exists y)(y \in x)]$

Pairing
$(\forall\varphi)(\forall\psi)(\exists x)(\exists y)[(x \in \varphi\,\&\,y \in \psi) \supset (\exists\gamma)(\varphi \in \gamma\,\&\,\psi \in \gamma)]$

Union
$(\forall\varphi)(\exists\psi)(\exists x)(\forall y)[(x \in \varphi\,\&\,y \in x) \supset y \in \psi]$

Power Set
$(\forall\varphi)(\exists\psi)(\forall x)(x \in \psi \equiv x \subseteq \varphi)$

Infinity
$(\exists\varphi)[\varnothing \in \varphi\,\&\,(\forall x)(x \in \varphi \supset x \cup \{x\} \in \varphi)]$

Choice
$(\forall\varphi)(\exists f)(D(f) = \{x | x \supseteq \varphi\,\&\,x \neq \varnothing\}\,\&\,(\forall\psi)[(\psi \subseteq \varphi \& \psi \neq \varnothing) \supset f(\psi) \in \psi])$

(*Replacement* and *Restriction* axioms are excluded as optional to set theoretical semantics.)

Comprehension and Free Assumption

The unrestricted set determination principle associates with every exclusively nuclear property of objects a set in the Meinongian domain. The principle is philosophically justified as the set theory counterpart of the *Annahmen* or unrestricted free assumption thesis in Meinongian object theory. If assumption is free, then it can freely postulate sets as objects of thought corresponding to any condition on objects. The naive determination of sets is standardly subject to set theory paradoxes involving self-non-membership diagonalizations. But these are avoided in Meinongian set theory and logic by restrictions on set abstraction. The proof of this aspect of the logic is beyond the scope of the present exposition, but is explained elsewhere.[9]

The naive comprehension principle or unrestricted determination of sets, and the fact that Meinongian domain D is itself an object, implies that D as the set of all objects is also (harmlessly) one of the objects and sets of objects in D. The set of all sets in D, the power set of $D, \mathcal{P}(D)$, is itself an object, and so must also belong to D. This entails by comprehension and the power set axiom that D contains its own power set; the power set of D is therefore a member and subset of $D, \mathcal{P}(D) \in D \,\&\, \mathcal{P}(D) \subseteq D$.

6.3 Intensional Set Theory Semantics

The predicate semantics for the object language of the logic can now be set theoretically defined. The function I is an interpretation that assigns objects as elements of the domain to items of syntax, terms, functors, predicates, descriptors, and abstracts. To each constant 'o_i' of O, I assigns an object in D. $I(\text{`}o_i\text{'}) = o_i$ indicates that $I$ assigns object $o_i$ to object constant '$o_i$'. To each $n$-ary functor or function symbol '$f_i^n$' of $O$, $I$ assigns an $n$-ary function from $D$ to $D$, $I(\text{`}f_i^n\text{'}) = f_i^n$. To each n-ary predicate '$P^n(!)_i$' of O, I assigns an n-ary property $P^n(!)_i, I(\text{`}P^n(!)_i\text{'}) = P^n(!)_i$. More precisely, $I$ assigns to each $n$-ary predicate '$P^n(!)_i$' of $O$ a set of ordered $n$-tuples of the objects in $D$, such that $I(\text{`}P^n(!)_i\text{'}) = \{\langle o_{i_1}, \ldots, o_{i_n}\rangle, \langle o_{j_1}, \ldots, o_{j_n}\rangle, \ldots\}$. Similarly for '$P^n(!)$'.

This provides an intensional counterpart of the identity of a property with the existent and nonexistent objects in the *intensions* of corresponding predicates, similar and in a way parallel to the identity of a property with the existent objects in the *extensions* of corresponding predicates in standard extensional semantics. The intension of a predicate is non-Fregean, though indirectly related to Frege's property 'senses' by virtue of the identity conditions for objects, determined by their *Soseine*, or associated unordered sets of nuclear constitutive properties.

Function V is a truth valuation on the well-formed formulas of O, which gives a unique truth value for every proposition as interpreted under I for D. The role of

[9] See Jacquette, *Meinongian Logic*, especially pp. 158-9, 174–86.

truth valuation function V for the propositional fragment of the logic has already been described. The valuations for predicate or quantificational theory can be explained by means of $I(\text{`}P^n(!)_i\text{'}), I(\text{`}\overline{P^n}(!)_i\text{'})$, and $I(\text{`}o_i\text{'})$, as previously defined.

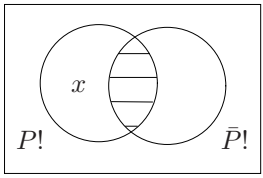
1. $V(P^n(!)o_1, \ldots, o_n) = T \equiv \langle I(\text{`}o_1\text{'}), \ldots, I(\text{`}o_n\text{'})\rangle \in I(\text{`}P^n(!)\text{'})$. $V(P^n(!)o_1, \ldots, o_n) = F \equiv \langle I(\text{`}o_1\text{'}), \ldots, I(\text{`}o_n\text{'})\rangle \notin I(\text{`}P^n(!)\text{'}).V(P^n o_1 \ldots o_n) = U \equiv -(\langle I(\text{`}o_1\text{'}), \ldots, I(\text{`}o_n\text{'})\rangle \in I(\text{`}P^n\text{'}) \vee \langle(\text{`}o_1\text{'}), \ldots, I(\text{`}o_n\text{'})\rangle \notin I(\text{`}P^n\text{'}))$ (for nuclear properties only).

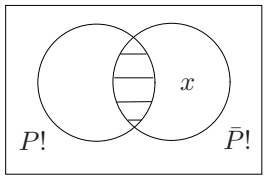
2. $V(\overline{P^n}(!)o_1 \ldots o_n) = T \equiv \langle(\text{`}o_1\text{'}), \ldots, I(\text{`}o_n\text{'})\rangle \in I(\text{`}\overline{P^n}(!)\text{'})$. $V(\overline{P^n}(!)o_1 \ldots o_n) = F \equiv \langle I(\text{`}o_1\text{'}), \ldots, I(\text{`}o_n\text{'})\rangle \notin I(\text{`}\overline{P^n}(!)\text{'}).V(P^n o_1 \ldots o_n) = U \equiv -(\langle I(\text{`}o_1\text{'}), \ldots, I(\text{`}o_n\text{'})\rangle \in I(\text{`}\overline{P^n}\text{'}) \,\&\, \langle I(\text{`}o_1\text{'}), \ldots, I(\text{`}o_n\text{'})\rangle \notin I(\text{`}P^n\text{'}))$ (for nuclear properties only).

3. $V(P^n o_1 \ldots o_n) = U \equiv (V(-P^n o_1 \ldots o_n) = U \,\&\, V(\overline{P^n} o_1 \ldots o_n) = U \,\&\, V(-\overline{P^n} o_1 \ldots o_n) = U)$ (for nuclear properties only).

The valuations for nuclear predications are diagrammed below. The extranuclear predications are classical, so their corresponding diagrams are standard with closed intersections that permit no membership outside the interlocking circles.

Extranuclear Predications

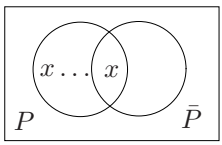


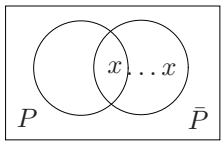


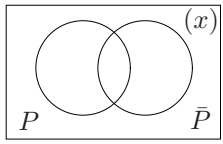
$V(P^n! o_1 \ldots o_n) = T$
$V(-P^n! o_1 \ldots o_n) = F$
$V(\overline{P^n}! o_1 \ldots o_n) = F$
$V(-\overline{P^n}! o_1 \ldots o_n) = T$

$V(P^n! o_1 \ldots o_n) = T$
$V(-P^n! o_1 \ldots o_n) = T$
$V(\overline{P^n}! o_1 \ldots o_n) = F$
$V(-\overline{P^n}! o_1 \ldots o_n) = F$

Nuclear Predications



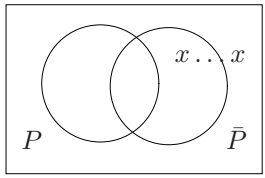


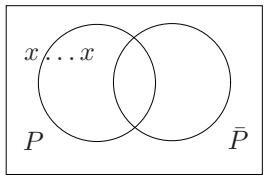


$V(P^n o_1 \ldots o_n) = T$
$V(-P^n o_1 \ldots o_n) = F$

$V(P^n o_1 \ldots o_n) = T$
$V(-P^n o_1 \ldots o_n) = F$

$V(P^n o_1 \ldots o_n) = U$
$V(\overline{P^n} o_1 \ldots o_n) = U$
$V(-P^n o_1 \ldots o_n) = U$
$V(-\overline{P^n} o_1 \ldots o_n) = U$





$V(P^n o_1 \ldots o_n) = F$ $\qquad$ $V(\overline{P^n} o_1 \ldots o_n) = F$

$V(-P^n o_1 \ldots o_n) = T$ $\qquad$ $V(-\overline{P^n} o_1 \ldots o_n) = T$

The diagrams for nuclear predications distinguish intensions of property P and its complement $\bar{P}$. Since some impossible objects have both property P and its complement in their *Soseine*, the spheres that represent the intensions of properties P and $\bar{P}$, or, the set of all existent and nonexistent objects with properties P or $\bar{P}$, are interlocking or overlapping. Unlike the previous diagram for the entire domain of O, in these diagrams objects can occur outside both spheres, to allow for incomplete objects lacking property P and its complement. The token 'x' indicates the location, inclusion or exclusion, of the ordered n-tuple of objects in $\langle o_1, \ldots, o_n \rangle$ in or from or with respect to the intensions of predicates 'P' and '$\bar{P}$'. The use of '$x \ldots x$', as in Venn diagrams, indicates that $\langle o_1, \ldots, o_n \rangle$ may be in either of the intensions where either 'x' in '$x \ldots x$' occurs. This enables a single diagram to represent several possibilities for the inclusion or exclusion of objects in or from or with respect to the intensions of predicates with the same truth valuational consequences. In the third diagram, '(x)' indicates that $\langle o_1, \ldots, o_n \rangle$ is neither included in nor excluded from either the intension of nuclear predicate 'P' or the intension of its complement '$\bar{P}$'. This means that it is undetermined with respect to inclusion in or exclusion from the intensions of either predicate or complement, or, in other words, from the sets of existent and nonexistent objects with the corresponding properties.

4. $V((\forall x)p) = T$ just in case for every truth valuation in $\langle D, I, V \rangle, V(p) = T$. $V((\forall x)p = F$ just in case there is at least one truth valuation in $\langle D, I, V \rangle$ such that $V(p) = F$. $V((\forall x)p) = U$ just in case there is at least one truth valuation in $\langle D, I, V \rangle$ such that $V(p) = U$, and there is no truth valuation in $\langle D, I, V \rangle$ such that $V(p) = F$.

There is no formal contradiction in $\overline{P^n}(!)o_1 \ldots o_n \,\&\, - P^n(!)o_1 \ldots o_n$, $-P^n o_1 \ldots o_n \,\&\, - P^n o_1 \ldots o_n$, or $P^n o_1 \ldots o_n \,\&\, - \overline{P^n} o_1 \ldots o_n$, just as there is no formal contradiction in $P^n o_1 \ldots o_n \& P^n o_1 \ldots o_n$. By the semantics of sentence negation and predicate complementation, it is invalid to infer either conjunct from the other.

6.4 Validity

Validity of deductive inference is defined by cases. The valid inference patterns are:

$$\begin{array}{lcl} T & \Rightarrow & T \\ F & \Rightarrow & F \\ F & \Rightarrow & T \\ F & \Rightarrow & U \\ U & \Rightarrow & T \\ U & \Rightarrow & U \end{array}$$

The remaining possible constructible combinations of truth values yield formally invalid inference patterns.

$$\begin{array}{lcl} T & \not\Rightarrow & F \\ T & \not\Rightarrow & U \\ U & \not\Rightarrow & F \end{array}$$

A proposition p is also said to be valid, a theorem or tautology of the logic, $\vdash p$, if and only if, for every semantic model $\langle D, I, V\rangle, V(p) = T$.

Comparison with the truth matrix for the conditional shows an exact correspondence between valid inferences and true conditionals, and invalid inferences and false or undetermined conditionals. This makes it possible to prove a nonstandard version of the Deduction Theorem, offered below in Section 7.

7 CONSISTENCY AND COMPLETENESS

This is a consistency, Henkin-type strong and weak completeness proof, based on a version first extended to three-valued logic by Leblanc, Goldberg, and Weaver.[10]

7.1 Internal Determinacy

It is useful to begin by establishing an internal determinacy theorem for Meinongian logic.[11] Although some propositions of O are logically true, and others logically false, no propositions of O are logically undetermined. The truth valuations under which an undetermined proposition can validly be inferred prevent deduction of an undetermined proposition from true propositions. The semantics of O do not require that any truth functionally complex proposition be undetermined in truth value. If the undetermined truth valuation of a proposition can only come from outside the logic in a particular scientific or philosophical application, then

[10] Hugues Leblanc, H. Goldberg, and George Weaver, "A Strong Completeness Theorem for Three–Valued Logic: Part I", in Leblanc, *Existence, Truth, and Provability* (Albany: State University of New York Press, 1982), pp. 240–246; Leblanc, "A Strong Completeness Theorem for Three–Valued Logic: Part II", pp. 247–257. Leon Henkin, "The Completeness of the First–Order Functional Calculus", *Journal of Symbolic Logic*, 14, 1949, 159-166.

[11] Jacquette, "An Internal Determinacy Metatheorem for Lukasiewicz's *Aussagenkalküls*", *Bulletin of the Section of Logic*, 29, 2000, 115–124.

within the logic inference always carries from true to true, and never to undetermined, propositions. This preserves an exact semantic isomorphism with the inference conditions of classical logic and the classical deduction theorem.

Internal Determinacy Theorem for O
No proposition of O is logically undetermined.

Proof. By mathematical induction on the length of wffs in O. Let p be a wff of O.

1. If $p = (q \supset r)$, then by truth tables, $V(p) = U$ if and only if either $V(q) = T$ and $V(r) = U$ or $V(q) = U$ and $V(r) = F$. But then p cannot be logically undetermined unless q or r is logically undetermined.
2. If $p = -q$, then by truth tables, $V(p) = U$ if and only if $V(q) = U$. But then p cannot be logically undetermined unless q is logically undetermined.
3. If p is not truth functionally complex, then $V(p) = T, V(p) = F$, or $V(p) = U$. But then p cannot be logically undetermined. ■

It follows that no proposition of O is logically or internally undetermined. The truth valuation of a proposition of O as undetermined must come from outside the logic. Undetermined nuclear predications of certain properties to incomplete nonexistent objects can only be introduced as substitution instances for propositional compoments of axioms and theorems in extralogical applications.

7.2 Propositional Object Theory Logic

DEFINITIONS

1. Let O' be the propositional fragment of O. A set ξ of wffs of O' is *syntactically consistent* if there is no wff p of O' such that both p and $-p$ are provable from ξ. Set ξ is *syntactically inconsistent* if there is a wff p of O' such that both p and $-p$ are deducible from ξ. As explained in the section on 'Inference Structures', to be deducible in O' means to be derivable in a sequence of wffs by the logical axioms, definitions, nonlogical axioms, and inference principles of O'. Set ξ is *maximally consistent* if (i) ξ is syntactically consistent; (ii) $\xi \vdash p$ for any wff p of O' such that $\xi \cup \{p\}$ is syntactically consistent.

2. Set ξ of wffs of O' is *semantically consistent* if there is a nonstandard truth value assignment for which all members of ξ have truth value T. Set ξ *entails* a wff p of $O'(\xi \vDash p)$ if, under any truth value assignment in O', p has truth value T when every member of ξ has truth value T. Proposition p is *valid* ($\vDash p$ of $\varnothing \vDash p$) if, under any truth value assignment in O', p unconditionally has truth value T.

LEMMA 1.

1. *If $\xi \vdash p$, then $zxi' \vdash p$ for every superset ξ' of ξ. By definition. In particular, if $\vdash p$, then $\xi \vdash p$ (which, by validity semantics, holds for every set ξ of wffs of O').*

2. *If $\xi \vdash p$, then there is a finite subset ξ' of ξ such that $\xi' \vdash p$. By definition.*

3. *If $p \in \xi$, then $\xi \vdash p$.*

4. *If $\xi \vdash p$ and $\xi \vdash (p \supset q)$, then $\xi \vdash q$.*

5. *If $\xi \cup \{p\} \vdash q$, then $\xi \vdash (p \supset (p \supset q))$.*

Proof. Assume that the wff sequence made up of $r_1, r_2, r_3, \ldots, r_n$ constitutes a proof of q from $\xi \cup \{p\}$. By mathematical induction on i it is shown that $\xi \vdash (p \supset (p \supset r_i))$ for each i from $1 - n$, and hence in particular that $\xi \vdash (p \supset (p \supset q))$. There are three cases to consider.

Case 1. r_i is an axiom or member of $\xi (r_i \in \xi)$.
Then $\xi \vdash r_i$, by (1) or (3). But $\xi \vdash (r_i \supset (p \supset r_i))$ by (1). Hence, $\xi \vdash (p \supset r_i)$ by (4). But $\xi \vdash ((p \supset r_i) \supset (p \supset (p \supset r_i)))$ by (1). Hence, $\xi \vdash (p \supset (p \supset r_i))$ by (4). In particular, then, $\xi \vdash (p \supset (p \supset q))$.

Case 2. $r_i = p$.
Then $\xi \vdash (p \supset (p \supset r_i))$ by (1). In particular, then, $\xi \vdash (p \supset (p \supset q))$.

Case 3. r_i is derived from r_j and $r_j \supset r_i$.
Then $\xi \vdash (p \supset (p \supset r_i))$ and $\xi \vdash (p \supset (p \supset (r_j \supset r_i)))$, by the hypothesis of the induction. Hence $\xi \vdash (p \supset (p \supset r_i))$ by (3) and (4). In particular, then, $\xi \vdash (p \supset (p \supset q))$. ■

6. If ξ is syntactically inconsistent, then $\xi \vdash p$, for every wrff p of O'.

Proof. Assume that $\xi \vdash q$ and $\xi \vdash -q$ for some wff q of O'. Then by truth tables and (4), $\xi \vdash p$ for any wff p of O'. ■

7. ξ is syntactically inconsistent if and only if $\xi \vdash (p \,\&\, -p)$.

Proof. By truth tables and internal determinacy, $\xi \vdash (p \supset p)$ for any wff p. Hence, if $\xi \vdash (p \,\&\, -p)$, then by truth tables and the definition, ξ is syntactically inconsistent. The converse follows by (6). ■

8. If $\xi \cup \{p\}$ is syntactically inconsistent, then $\xi \vdash -p$.

Proof. Assume that $\xi \cup \{p\}$ is syntactically inconsistent. Then $\xi \cup \{p\} \vdash -p$ by (6). Thus, $\xi \vdash (p \supset -p)$ by (5). Thus, $\xi \vdash -p$ by truth tables and (4). ■

9. If $\xi \cup \{-p\}$ is syntactically inconsistent, then $\xi \vdash p$.

Proof. From (8), (1), and (4). ■

LEMMA 2. *If any set ξ of wffs of O' is syntactically consistent, then ξ is also semantically consistent.*

The proof involves the assumption that ξ is syntactically consistent, and the construction of a ξ extension into a two-valued superset $\xi*$. The members of $\xi*$, and hence of ξ, are shown to have truth value T under a special assignment that constitutes a Henkin model for ξ in O'. The construction of $\xi*$ requires several steps. Let ξ_0 be ξ. Assume that the wffs of O' are alphabetically well-ordered in some unambiguous enumeration, and consider p_i to be, for each $i \geq 1$, the ith wff of O'. Then ξ_i is $\xi_{i-1} \cup \{p_i\}$ if $\xi_{i-1} \cup \{p_i\}$ is syntactically consistent, and otherwise ξ_i is ξ_{i-1}. Finally, $\xi* = S(\xi_i)$.

It is shown that:

1. *$\xi*$ is syntactically consistent.*
2. *$\xi*$ is maximally consistent.*

Proof.

1. Assume for purposes of indirect proof that $\xi*$ is syntactically inconsistent. Then by (7) and (2) of Lemma 1 at least one finite subset ξ' of $\xi*$ is syntactically inconsistent. But ξ' must either be a subset of ξ_0, ξ_1, ξ_2, ξ_3,..., each of which is syntactically consistent by condition of membership in $\xi*$.
2. Assume that $\xi* \nvdash p_i$, where p_i is the ith wff of O' in the same unambiguous alphabetical ordering as before. Then by (3) of Lemma 1, $p_i \notin \xi*$. Thus, $p_i \notin \xi_i$, and $\xi_{i-1} \cup \{p_i\}$ is syntactically inconsistent. Hence, by (7) and (1) of Lemma 1, $\xi* \cup \{p_i\}$ is also syntactically inconsistent. ■

THEOREM 3 Consistency Theorem for O'. *Let there be a truth value assignment in $O', V_{O'}$, that assigns to each propositional variable 'p_i' of O' the value T if $\xi* \vdash p_i$ (so that by the syntactical consistency of $\xi*, \xi* \nvdash -p_i$); the value F if $\xi* \vdash -p_i$ (so that by the syntactical consistency of $\xi*, \xi* \nvdash p_i$); and otherwise assigns p_i the undetermined truth value U.*

For any wff p_i of O' it is shown that:

1. *If $\xi* \vdash p_i$ (where $\xi* \nvdash -p_i$), $V(p_i) = T$.*
2. *If $\xi* \vdash -p_i$ (where $\xi* \nvdash p_i$), $V(p_i) = F$.*
3. *If $\xi* \nvdash p_i$ and $\xi* \nvdash -p_i$, $V(p_i) = U$.*

Proof. The proof is by mathematical induction on the length of p_i. The length $L(p_i)$ of a propositional variable 'p_i' is 1. The length $L(-p_i)$ of a negation '$-p_i$' is $L(p_i) + 1$. The length $L(p_i \supset p_j)$ of a conditional '$p_i \supset p_j$' is $L(p_i) + L(p_j) + 1$. It is assumed as confirmable by truth table analysis and the recursive inductive

definition of the wffs of O' that any wff of O' can be truth functionally reduced to equivalent form in the primitive notation, involving only negation and the conditional.

Basis: $L = 1$.
Here p_i is a propositional variable. The proof follows from the definition of $V_{O'}$.
Inductive step: $L > 1$.
There are two case to consider.

Case 1. $p_i = -q$

1. Assume that $\xi* \vdash -q$. Then it is false that $\xi* \vdash q$. Hence, by the hypothesis of the induction, $V(q) = T$ and $V(-q) = F$.
2. Assume that $\xi* \vdash -q$. Then by truth tables and (4) of Lemma 1, $\xi* \vdash q$. Hence, by the hypothesis of the induction, $V(q) = T$, and $V(-q) = F$.
3. Assume that neither $\xi* \vdash -q$ nor $\xi* \vdash -q$. If q were provable from $\xi*$, then by truth tables and (4) of Lemma 1, $-q$ would also be provable from $\xi*$. Hence, neither $\xi* \vdash q$ nor $\xi* \vdash -q$. Thus, by the hypothesis of the induction, $V(q) = U$ and $V(-q) = U$.

Case 2. $p_i = (q \supset r)$

1. Assume that $\xi* \vdash (q \supset r)$. If $\xi* \vdash q$, then $V(q) = T$ by the hypothesis. If $\xi* \vdash r$, then $V(r) = T$, again by the hypothesis of the induction. Thus, if $V(q) = T$ and $V(r) = T$, then $V(q \supset r) = T$.
2. Assume that $\xi* \vdash -(q \supset r)$. Then by truth tables $\xi* \vdash q$ and $\xi* \vdash -r$. Thus, by the hypothesis of the induction, $V(q) = T$ and $V(r) = F$. Therefore $V(q \supset r) = F$.
3. Assume that neither $\xi* \vdash (q \supset r)$ nor $\xi* \vdash -(q \supset r)$. Then neither $V(r) = T$ nor $V(q) = F$. For by the hypothesis of the induction, $q, r, -q$, or $-r$ would then be deducible from $\xi*$, and thus, by the above, so would $q \supset r$ or $-(q \supset r)$, contrary to the assumption. Suppose that $V(q) = T$. Then $V(r) \neq T$ and $V(r) \neq F$. For then by the hypothesis of the induction, $q \supset r$ or $-(q \supset r)$ would be deducible from $\xi*$, contrary to the assumption. $V(q) \neq T$ and $V(q) \neq F$, so $V(q) = U$. But if $V(q) = U, V(r) \neq T$, for then by truth tables and the hypothesis of the induction, $q \supset r$ would be deducible from $\xi*$, contrary to the assumption. Then $V(r) \neq U$, for then by truth tables and the hypothesis of the induction, $q \supset r$ would be deducible from $\xi*$, again contrary to the assumption. Therefore, $V(q) = T$ and $V(r) = U$ or $V(q) = U$ and $V(r) = F$; and in either case, $V(q \supset r) = U$ and $V(-(q \supset r)) = U$.

Since every member of ξ is a member of $\xi*$, every member of ξ is provable from $\xi*$ by (3) of Lemma 1. This means that every member of ξ has truth value T under truth value assignment $V_{O'}$. Meinongian propositional object theory logic O' therefore has a Henkin model. It follows that:

If ξ is syntactically consistent, then ξ is semantically consistent. ■

THEOREM 4 Strong Completeness Theorem for O'. *If* $\xi \vDash p$, *then* $\xi \vdash p$.

Proof. Assume that $\xi \vDash p$. Then $\xi \cup \{-p\}$ is semantically inconsistent. Hence, by Lemma 2, $\xi \cup \{-p\}$ is syntactically inconsistent. From this and (9) of Lemma 1, it follows that $\xi \vdash p$. ■

THEOREM 5 Weak Completeness Theorem for O'. *If* $\vDash p$, *then* $\vdash p$.

Proof. From strong completeness, where $\xi = \varnothing$. ■

THEOREM 6 Compactness Theorem for O'. *If every finite subset of* ξ *is semantically consistent, then* ξ *is semantically consistent.*

Proof. If ξ is syntactically inconsistent, then by (7) of Lemma 1, there is a proposition p of O' such that $\xi \vdash (p\&-p)$. But then ξ is semantically inconsistent, since by the hypothesis of the induction in the proof of Lemma 2, $V(p) = T$ and $V(-p) = F$. Thus, if ξ is syntactically inconsistent, then ξ is semantically inconsistent. This implies by contraposition that if ξ is semantically consistent, then ξ is syntactically consistent (the converse of Lemma 2). It follows from this and (2) and (7) of Lemma 1 that if every finite subset of x is semantically consistent, then x is syntactically consistent. This, with Lemma 2, entails the theorem. ■

7.3 Predicate Object Theory Logic

DEFINITIONS

1. A set ξ of wffs of O is *syntactically decidable* with respect to truth value if there is a truth value assignment in O under which all members of x are assigned truth value T.

2. A set ξ of wffs of O is *semantically consistent* if either ξ or some set isomorphic to ξ is syntactically decidable with respect to truth value. Set ξ *entails* a wff p if $\xi \cup \{-p\}$ is semantically inconsistent.

Let ξ be a syntactically consistent and infinitely extendible set of wffs of O. Set ξ is extended into another set $\xi**$, then $\xi**$ is extended into $\xi***$, and $\xi***$ extended into a set $\xi*$, like that of the propositional metatheory, to which predicate are added. It is shown that there is a truth value assignment according to which

every wff in set ξ receives truth value T. Set $\xi{**}$ is defined as follows. Let ξ^0 be ξ. Assume that the quantified wffs of O are ordered alphabetically so that $(\forall x_i)p_i$ is the ith wff of O for each $i \geq 1$. Take ξ^i to be $\xi^{i-1} \cup \{p_i(o_i/x_i) \supset (\forall x_i)p_i\}$, where '$o_i$' designates any object of the domain not designated in ξ^{i-1} such that $(\forall x_i)p_i$. Let ξ^{**} be the union set $\xi^0 \cup \xi^1 \cup \xi^2 \cup \xi^3 \cup \ldots \cup \ldots$. Set $\xi{***}$ is defined as the extension of $\xi{**}$. Let $\xi^0_m = \xi{**}$. Set $\xi{*}$ is defined as the extension of $\xi{***}$. Let $\xi_0 = \xi{***}$. Proposition p_i is the ith wff of O in the expanded ordering of wffs of $\xi{***}$. Let $\xi_i = \xi_{i-1} \cup \{p_i\}$ if $\xi_{i-1} \cup \{p_i\}$ is syntactically consistent, and otherwise take ξ_i to be ξ_{i-1}. Let $\xi{*}$ be the union set $\xi_0 \cup \xi_1 \cup \xi_2 \cup \xi_3 \cup \ldots \cup \ldots$.

It is shown that:

1. $\xi{***}$ is syntactically consistent.
2. $\xi{**}$ is syntactically consistent.
3. $\xi{*}$ is syntactically consistent.
4. $\xi{*}$ is maximally consistent.

Proof.

1. Assume for purposes of indirect proof that $\xi{***}$ is syntactically inconsistent. Then by (7) and (2) of Lemma I, at least one finite subset $\xi * *'$ of $\xi{***}$ is syntactically inconsistent. But $\xi * *'$ must either be $\xi^0_m, \xi^1_m, \xi^2_m, \xi^3_m, \ldots$, each of which is syntactically consistent by membership conditions for the construction of $\xi{***}$.

2. Assume that ξ^i is syntactically inconsistent. Then by (8) of Lemma 1, $-(p_i(o_i/x_i) \supset (\forall x_i)p_i)$ is provable from ξ^{i-1}. Let $x_{i'}$ be the ith variable ranging over individuals of O not designated in $(\forall x_i)p_i$. Then $\xi^{i-1} \vdash (\forall x_{i'}) - (p_i(x_{i'}/x_i) \supset (\forall x_i)p_i)$. From this it follows that:

$$\xi^{i-1} \vdash (\forall x_{i'}) -(p_i(x_{i'}/x_i) \supset (\forall x_i)p_i) \supset -(\exists x_{i'})(p_i(x_{i'}/x_i) \supset (\forall x_i)p_i)$$
$$\xi^{i-1} \vdash -(\exists x_{i'})(p_i(x_{i'}/x_i) \supset (\forall x_i)p_i)$$
$$\xi^{i-1} \vdash (\exists x_{i'})(p_i(x_{i'}/x_i) \supset (\forall x_i)p_i) \supset -(\exists x_{i'})(p_i(x_{i'}/x_i) \supset (\forall x_i)p_i)$$
$$\xi^{i-1} \vdash (\exists x_{i'})(p_i(x_{i'}/x_i) \supset (\forall x_i)p_i)$$

Thus, ξ^{i-1} is syntactically inconsistent. But ξ^i is syntactically consistent if ξ^{i-1} is. By hypothesis, ξ^0 is syntactically consistent. Thus, each of ξ^0, ξ^1, ξ^2, ξ^3,... is syntactically consistent. Hence, by (1) and (2) above, $\xi{**}$ is syntactically consistent.

3. Assume that $\xi{*}$ is syntactically inconsistent. Then by (7) and (2) of Lemma 1, at least one finite subset $\xi{***}'$ of $\xi{*}$ is syntactically inconsistent. But $\xi{***}'$ must either be a subset of ξ_0, ξ_1, ξ_2, ξ_3,..., each of which is syntactically consistent by membership conditions for construction of $\xi{*}$.

4. Assume it is not the case that $\xi* \vdash p_i$, where p_i is the ith wff of O in the unique ordering of $\xi*$. Then by (3) above, $p_i \notin \xi*$. Hence, $p_i \notin \xi**'$. Thus, $\xi** \cup \{p_i\}$ is syntactically inconsistent, and by (7) and (1) of Lemma 1, $\xi* \cup \{p_i\}$ is also syntactically inconsistent. Thus, $\xi*$ is maximally consistent. ■

Let V_O be a truth value assignment in O that assigns to each propositional variable 'p_i' of O the truth value T if $\xi* \vdash p_i$, F if $\xi* \vdash -p_i$, and U if neither $\xi* \vdash p_i$ nor $\xi* \vdash -p_i$. It is proved for any wff p_i of O:

1. If $\xi* \vdash p_i$, then $V(p_i) = T$.
2. If $\xi* \vdash -p_i$, then $V(p_i) = F$.
3. If neither $\xi* \vdash p_i$ nor $\xi* \vdash -p_i$, then $V(p_i) = U$.

Proof. The proof is by mathematical induction on the length of wffs. The length L of a wff, already defined for propositional formulas, is now defined for predicate formulas. The length $L((\forall x)p_i)$ of a universally quantified formula '$(\forall x)p_i$' is $L(p_i(o_i/x)) + 1$, where '$(p_i(o_i/x)$' indicates as before the replacement of object term 'o_i' for all occurrences of 'x' in p_i. The subtheories of lambda abstraction and definite description do not affect induction on length of wffs in O, since they only reductively introduce new kinds of object- and property-designating terms.

Basis: $L = 1$
Here p_i is a propositional variable. The proof follows from the definition of V_O.

Inductive step: $L > 1$
There are three cases to consider.

Case 1. $p_i = -q$
Proof as in propositional metatheory.

Case 2. $p_i = (q \supset r)$
Proof as in propositional metatheory.

Case 3. $p_i = (\forall x)q$

1. Assume that $\xi* \vdash (\forall x)q$. Then $\xi* \vdash q(o_i/x)$ for every object o_i of O. Thus, by the hypothesis of the induction, $V(q(o_i/x)) = T$ for every o_i, and hence, $V((\forall x)q) = T$.
2. Assume that $\xi* \vdash -(\forall x)q$. Let o_i be the ith object of O, such that $q(o_i/x) \supset (\forall x)q$. Hence, $\xi* \vdash -(\forall x)q \supset -q(o_i/x)$. Thus, $\xi* \vdash -q(o_i/x)$, and, by the hypothesis of the induction, $V(q(o_i/x)) = F$, so that $V((\forall x)q) = F$.
3. Assume that neither $\xi* \vdash (\forall x)q$ nor $\xi* \vdash -(\forall x)q$. If $V(q(o_i/x)) = F$ for any object o_i of O, then by the hypothesis of the induction, $-q(o_i/x)$ would be provable from $\xi*$ for some o_i, and hence,

$-(\forall x)q$ would also be provable from $\xi*$ for some o_i, contrary to the assumption. But if $V(q(o_i/x)) = T$ for every object o_i of O, then by the hypothesis of the induction, $q(o_i/x)$ would be provable from $\xi*$ for every o_i. But $q(o_i/x) \supset (\forall)q$ belongs to $\xi*$, and by (3) above is provable from $\xi*$ for at least some object o_i of O. Thus, if $V(q(o_i/x)) = T$ for every object o_i of O, then by (4), $(\forall x)q$ would be provable from $\xi*$, also contrary to the assumption. Hence, $V((\forall x)q) = U$.

Since every member of ξ is a member of $\xi*$, every member of $\xi*$ is deducible from $\xi*$ by (3) of Lemma 1. This means that every member of ξ has truth value T under truth value assignment V_O. Meinongian predicate object theory O therefore has a Henkin model.

It follows that:

If ξ is syntactically consistent and infinitely extendible, then ξ is syntactically decidable with respect to truth value, and hence semantically consistent. ■

THEOREM 7 Strong Completeness Theorem for O. *If* $\xi \vDash p$, *then* $\xi \vdash p$.

Proof. Assume that $\xi \vDash p$. Then $\xi \cup \{-p\}$ is semantically inconsistent. Hence, by Lemma 2, $\xi \cup \{-p\}$ is syntactically inconsistent. From this and (9) of Lemma 1 it follows that $\xi \vdash p$. ■

Weak Completeness Proof for O If $\vDash p$, then $\vdash p$.

Proof. From strong completeness, where $\xi = \varnothing$. ■

8 DEDUCTION THEOREM FOR O

If p is a wff of O, then for every wff q of O, under internal determinacy, $\vdash (p \supset q)$ if and only if q is a theorem of O/p (where O/p is the theory obtained from O by adding p as a nonlogical axiom).

Proof.

1. If $\vdash (p \supset q)$, then p and $p \supset q$ are theorems of O/p. Then q is a theorem of O/p.

2. If q is an axiom of O/p, then if $q = p$, then $(p \supset q) = (p \supset p)$, which is a theorem of O, and therefore of O/p.

3. If q is an axiom of O/p, but $q \neq p$, then q is an axiom of O, and not just of O/p. Then q is a theorem of O. It follows that $\vdash (p \supset q)$.

4. If $(r_1, \ldots, r_n) \vdash q$, then $(p \supset r_1, \ldots, p \supset r_n) \vdash (p \supset q)$. By the hypothesis of the induction, $\vdash (p \supset r_1), \ldots, \vdash (p \supset r_n)$. It follows that $\vdash (p \supset q)$. ■

The conditional version of the deduction theorem holds for internally determinate Meinongian object theory logic, and for its bivalent extranuclear fragment, even in scientific and philosophical applications involving undetermined nuclear predications to incomplete objects.[12]

9 CONSISTENCY AND FREE ASSUMPTION

The *Annahmen* or unrestricted freedom of assumption thesis in Meinongian logic underwrites a powerful comprehension principle for the object theory semantic domain. It is not so powerful, however, as to permit the ingress of outright logical inconsistency or logically inconsistent Meinongian objects.

From a phenomenological standpoint, the assumption that there is an object with both nuclear properties P and not-P is most naturally formalized as $(\exists x)(Px\,\&\,\overline{Px})$, and as such standardly comprehends an impossible but by no means logically inconsistent Meinongian object o_i such that $Po_i\,\&\,\overline{Po_i}$ or $\lambda x[Px\,\&\,\overline{Px}]o_i$. Here the object is freely assumed to have the above-named metaphysically impossible but logically consistent nuclear property or conjunction of nuclear properties in its *Sosein*.

Logical inconsistency is avoided where free assumption and the comprehension principle might otherwise be thought to introduce it in exploiting the syntactical resources of the logic, by the restriction of the *Annahmen* thesis to the determination of objects exclusively by their nuclear properties. The assumption that there is an object such that it has nuclear property P and it is *not the case that* it has nuclear property P, $(\exists x)(Px\&-Px)$, does not comprehend a logically inconsistent Meinongian object. Although 'P' and '$\bar{P}$' are (nuclear) property terms, '$-P$', according to the formation rules, is not a term at all, but an ill-formed symbol combination that represents no nuclear or extranuclear property. The logical inconsistency abstract '$\lambda x[Px\,\&\,-Px]$', on the other hand, is a term, but for a decidedly *extranuclear* property to which the logic's comprehension principle via free assumption does not apply. The abstract is correctly written '$\lambda x[Px\&-Px]!$', as in the assumption $(\exists y)(\lambda x[Px\,\&\,-Px]!y)$, since it evidently satisfies criterion (C2) for extranuclear properties:

$$(\forall y)(\overline{\lambda x[Px\,\&\,-Px]}!y \equiv -(\lambda x[Px\,\&\,-Px]!y))$$

The property of being logically inconsistent, like the property of being possible, impossible, or logically consistent, is intuitively extranuclear rather than nuclear. That an object has the complement of the property of being logically inconsistent (or, is logically consistent) is logically equivalent to the object's not being logically inconsistent or to it not being the case that the object is logically inconsistent.[13]

[12]This formulation of the deduction theorem is adapted from J. R. Shoenfield, *Mathematical Logic* (Reading: Addison–Wesley Publishing Company, 1967), p. 33.

[13]See Jacquette, "Object Theory Foundations for Intensional Logic", *Acta Analytica*, 13, 1995, pp. 33–63.

10 APPLICATION: IMPLEXIVE BEING, SO-BEING, AND NONBEING

The concept of implexive being in Meinong's later thought makes it possible to understand the relation between the *Soseine* of distinct intended objects. It also facilitates an extraordinary reduction in the Meinongian semantic domain, whereby it is in many respects more austere than such quasi-nominalistic ontologies as those of Quine.

Implexive being, appropriately enough, makes its first appearance in Meinong's text as an explanation of the metaphysics of universals. Meinong, as a reflection of his commitment to a Brentanian empiricism, accepts Berkeley's rejection of abstract general ideas as incomplete.[14] Berkeley and Hume, from their respective empiricist standpoints, seek to solve at most what might be called the epistemic problem of universals. They reject universals or abstract general ideas because of their essential incompleteness, and find it necessary thereafter only to account for what is popularly regarded as reasoning about abstract general ideas, given that in fact there are no such ideas.

Berkeley's ingenious solution, which Hume in *A Treatise of Human Nature* admires as "one of the greatest and most valuable discoveries that has been made of late in the republic of letters",[15] has two parts. The intuitive belief that we seem to reason about universals or abstract general ideas is first interpreted by the proposition that particular ideas are delegated in thought to represent any of a class or category of similar particular ideas. Then, apparently as revealed by a somewhat idealized introspection, Berkeley claims that if we wrongly try to draw inferences about the properties of all the ideas in a category from the peculiarities of the particular idea representing the others, then counterexamples that do not

[14]George Berkeley, *A Treatise on the Principles of Human Knowledge*, *The Works of George Berkeley Bishop of Cloyne*, edited by A.A. Luce and T.E. Jessop (London: Thomas Nelson & Sons, 1949), Vol. II, p. 45. Berkeley, *Three Dialogues Between Hylas and Philonous*, *Works*, II, pp. 192-194. David Hume, *An Enquiry Concerning Human Understanding* [1748], in *David Hume: Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, edited by L.A. Selby-Bigge, 3rd ed. revised by P.H. Nidditch (Oxford: The Clarendon Press, 1975), pp. 154-155: "An extension, that is neither tangible nor visible, cannot possibly be conceived: and a tangible or visible extension, which is neither hard nor soft, black nor white, is equally beyond the reach of human conception. Let any man try to conceive a triangle in general, which is neither *Isosceles* nor *Scalenum*, nor has any particular length or proportion of sides; and he will soon perceive the absurdity of all the scholastic notions with regard to abstraction and general ideas." Compare Meinong, *Über Möglichkeit und Wahrscheinlichkeit*, p. 178: "Besonders geeignet sind vielmehr Begriffsgegenstände, wie uns deren etwa durch Definitionen gegeben werden. Das Dreick z.B., darin hatte der sicher nicht überrationalistische Locke gegen Berkeley und gegen viele Spätere) am Ende doch recht, ist als solches weder gleichseitig noch gleichschenklig, weder rechtwinklig noch schiefwinklig, noch das Gegenteil davon: es ist in diesen Hinsichten und noch in vielen anderen eben unbestimmt. Gegenstände dieser Art stehen in deutlichen Gegensatz zu solchen, die, wie wir deren oben zuerst betrachtet haben, in bezug auf alle wie immer gearteten Gegenstände bestimmt sind. Man kann solche Gegenstände mit Recht vollständig bestimmte nennen, Blaues, Dreieck und ihresgleichen dagegen unvollständig bestimmte."

[15]Hume, *A Treatise of Human Nature*, [1739-40], edited by Selby-Bigge, second edition revised by Nidditch (Oxford: The Clarendon Press, 1978), p. 17.

share the same property will ordinarily occur at once to thought, rushing in to challenge any false unwarranted generalizations.[16]

This explains or explains away what is called 'reasoning about universals'. A phenomenologically plausible nominalism accounts for abstract thinking without adding universals or general abstract entities to the ontology. The solution nevertheless leaves unanswered a cluster of important metaphysical problems. These include the nagging suspicion that there must be something more in common between the particular triangles in the category 'the triangle', by virtue of which particular triangles are rightly judged to belong to the same category. The suggestion that all particulars of a certain sort are triangles solely because it is convenient or serves pragmatic interests to classify them together or give them all the same general term is thought by many to be philosophically unsatisfying. We may feel strongly that there must be more to it than that, even if we share empiricism's distrust of predicationally incomplete universals or abstract general ideas as existent or subsistent entities. Arguably, the best approach might then be to advance some version of Aristotle's thesis that the forms, definitions, universals, or secondary substances inhere or are somehow contained inherently in particular existent primary substances, and that it is by virtue of such inherence of 'the triangle' in any particular existent triangle that all particular triangles belong together in the same category, to be rightly referred to by the same universal or abstract general concept term.

It is to resolve the metaphysical problem of universals, left over from the British empiricists' solution to the epistemic problem of universals after universals have been eliminated from ontology, that Meinong first puts forward the concept of implexive being and nonbeing. Meinong agrees with Berkeley and Hume that universals are essentially incomplete, and that therefore universals are neither existent nor subsistent. He begins to cultivate these insights with his very first historical research in philosophy, undertaken at Brentano's suggestion, leading to his *Habilitationsschrift* on Hume's nominalism. This work was published in 1878 as the *Hume-Studien I* in the *Sitzungsberichte der Wiener Akademie der Wissenschaften*, and was followed in 1882 by the sequel on Hume's theory of relations, the *Hume-Studien II*.[17] For Meinong, however, after he has elaborated

[16]Berkeley, *A Treatise on the Principles of Human Knowledge*, pp. 29-40; *Three Dialogues Between Hylas and Philonous*, pp. 192-197. Hume endorses Berkeley's theory of representative generalization in place of abstract generalization in *An Enquiry Concerning Human Understanding*, p. 158, n. 1: "...all general ideas are, in reality, particular ones, attached to a general term, which recalls, upon occasion, other particular ones, that resemble, in certain circumstances, the idea, present to the mind."

[17]Meinong completed his 1877 *Habilitationsschrift* on Hume's nominalism, undertaken on Brentano's recommendation, and appearing as the *Hume-Studien I* in 1878 in the *Sitzungsberichte der Wiener Akademie der Wissenschaften*, and followed by a sequel in 1882 on Hume's nominalist theory of relations, the *Hume-Studien II*. See Meinong, "Zur Gegenstandstheorie", in Raymund Schmidt, ed., *Die Philosophie der Gegenwart in Selbstdarstellungen* (Leipzig: Felix Meiner Verlag, 1921). An English translation of Meinong's *Hume-Studien I, II*, respectively, on Hume's nominalism and theory of relations (*Gesamtausgabe*, I, II), is given by Kenneth Frank Barber, *Meinong's 'Hume Studies': Translation and Commentary*, Ph.D. thesis, The University

the basic principles of *Gegenstandstheorie*, the fact that universals like 'blueness' or 'the triangle' are beingless incomplete objects means only that they belong to a special category of the extraontological semantic domain or *Außersein*. True to Brentano's empiricist methodology and Aristotelian metaphysics, Meinong tries to answer the question why all blue things are blue and all triangular things triangles, by explaining in object theory terms what we can speak of somewhat anachronistically as related to Aristotle's notion of the inherence of secondary substances in primary substances by the concept of the implection of beingless essentially incomplete universals in complete existent or subsistent particulars.[18]

Meinong explains the metaphysics of universals by classifying them as beingless incomplete objects implected in complete existent or subsistent objects. The concept of implection appears in this context in Meinong's *Über Möglichkeit und Wahrscheinlichkeit*, when he writes:

> Now, if we consider part-whole relations, we can characterize the relationship of the incomplete object in its entirety, not only in terms of what it is not, but also in terms of what it is; thereby in particular establishing the validity of what has often enough been conjectured. Without a doubt, 'the ball', for example, has an exact meaning only if this or that particular ball exists; and in this way the turn of phrase 'the ball exists in this or that particular ball' surely makes good sense as a trope or analogy. I venture similarly to say the same, by means of a special terminology, that 'the ball' is implected in my friend's billiard ball.[19]

Meinong offers a positive account of incomplete objects. He wants to move beyond the purely negative statement that incomplete objects are *other than* or that they are *not* complete objects. He sketches a theory in terms of a part-whole relation between incomplete objects as in some sense mereologically though non-physically contained in complete objects. Then he indicates that the mereological concept affords insight into what we can describe as the Aristotelian view that the universal ball in some sense exists and may exist in all and only the particular

of Iowa, 1966 (Ann Arbor: University Microfilms, Inc., 1967). See also, Barber, "Meinong's Hume Studies, Part I: Meinong's Nominalism; Part II: Meinong's Analysis of Relations", *Philosophy and Phenomenological Research*, 30, 1970, pp. 550-567; 31, 1971, pp. 564-584. See Jacquette, "Alexius Meinong 1853-1920", in *The School of Franz Brentano*, edited by Roberto Poli and Liliana Albertazzi (Dordrecht: Kluwer Academic Publishing, 1996), pp. 131-159.

[18] Aristotle, *Categories* 2^a10-16; *Metaphysics* 1017^b1-25.

[19] Meinong, *Über Möglichkeit und Wahrscheinlichkeit*, p. 211 (my translation). The original reads: 'Nun darf uns aber das Verhältnis des unvollständigen Gegenstandes zum vollständigen nicht nur hinsichtlich dessen interessieren, was es nicht ist, sondern auch hinsichtlich dessen, was es ist, zumal dabei zur Geltung kommen kann, was mutmaßlich oft genug eigentlich gemeint worden ist, wenn man zur Beschreibung der Sachlage die Relation das Teiles zum Ganzen heranzog. Ohne Zweifel bedeutet es nämlich doch etwas für "die Kugel", when "eine Kugel", genauer also, wenn diese oder jene bestimmte Kugel existiert, und als Tropus oder Analogie ist der Wendung "die Kugel existiert in dieser oder jene bestimmten Kugel" sicher ein guter Sinn beizulegen. Ich versuche der Gefahr, Ähnliches für gleich zu nehmen, durch besondere Benennung vorzubeugen, indem ich von "der Kugel" sage, sie sei in der Billiardkugel meines Freundes "Implektiert".'

balls in which it inheres or is instantiated. But what exactly is the relation of an incomplete object to the complete object in which it is implected? In what sense are we to imagine ballness, 'the ball' or the universal ball as existing, not in the usual way, but implexively, as having implexive being in Meinong's friend's particular actually existing billiard ball?

Meinong distinguishes between different objects by identity conditions based on individuating *Sein*-independent and *Sein*-indifferent *Soseine* consisting of distinct sets of constitutive properties.[20] To each unique set of such properties there is associated a unique Meinongian object. The same Leibnizian identity principles hold with respect to beingless as well as existent and subsistent entities. This makes it possible to understand Meinong's concept of implexive being, and of the mereological implection of incomplete beingless objects in complete existent or subsistent objects, in terms of set theoretical relations between the *Sosein* of the incomplete beingless object and the *Sosein* of the complete existent or subsistent object in which the implected object is implected.

The *Sosein* of my friend's billiard ball consists among many other things of this incompletely specified predicationally complete set of constitutive properties:

> *Sosein*(my friend's billiard ball) = {spherical; physical (spatiotemporal); black; marked with the number '8' in a circular white field; weighing x grams; carved from illegally imported ivory; used in playing billiards; belonging to my friend; . . . }

The *Sosein* of the universal ball or 'the ball' as a beingless incomplete object, implected in my friend's billiard ball, according to Meinong's concept (if for simplicity sake we leave out of account differently shaped 'balls' like rugby balls and American footballs), might be understood on the present interpretation as consisting of the following incompletely specified predicationally incomplete set of constitutive properties:

$$\textit{Sosein}(\text{'the ball'}) = \{\text{spherical; } \ldots \text{; physical (spatiotemporal)}\}$$

Meinong's concept of *Sein*-independent *Sein*-indifferent *Sosein* associates with every distinct abstract set of constitutive individuating properties a unique existent, subsistent, or beingless object. This makes it possible to interpret implection criteriologically as the inclusion of a proper subset of the constitutive properties of an implected object in the set of constitutive properties in which the implected

[20] Meinong, "Über Gegenstandstheorie", *Gesamtausgabe*, II, pp. 490-493. See also Roderick M. Chisholm, "Beyond Being and Non-Being", *Philosophical Studies*, 24, 1972, pp. 245-255. Reinhardt Grossmann, "Meinong's Doctrine of the *Aussersein* of the Pure Object", *Noûs*, 8, 1974, pp. 67-82. The independence of *Sosein* from *Sein* thesis is formulated by Meinong's student Ernst Mally, in his "Untersuchungen zur Gegenstandstheorie des Messens", in Meinong, ed., *Untersuchungen zur Gegenstandstheorie und Psychologie* (Leipzig: Verlag von Johann Ambrosius Barth, 1904), p. 127. Findlay, *Meinong's Theory of Objects and Values*, p. 44. Karel Lambert, "A Logical Interpretation of Meinong's Principle of Independence", *Topoi*, 1, 1982, pp. 87-96. Lambert, *Meinong and the Principle of Independence*. Nicholas Griffin, "The Independence of *Sosein* from *Sein*", *Grazer Philosophische Studien*, 9, 1979, pp. 23-34.

object is implected. We can proceed now lightly to formalize the intuitive set theoretical relations that obtain when an incomplete beingless Meinongian object is implected in a complete existent or subsistent object.

The concept of implection, like intuitive proper subset membership, is, in every interesting application, unidirectional. If object o_1 is properly implected in object o_2, then object o_2 is not implected in object o_1, just as when a set S_1 is a proper subset of set S_2, then S_2 is not a subset of S_1. If we wish to allow improper implection as a limiting case, in which objects are interimplected, then, by means of an obvious symbolization, we can say that if any incomplete object o_1 is properly implected in any complete object o_2, then o_2 is not (properly or improperly) implected in o_1. More abstractly than in Meinong's example, we can then explain implexive being and nonbeing in set theoretical terms as the proper subset membership of the elements of one *Sosein* set of constitutive properties of the implected object in the *Sosein* set of constitutive properties of the object in which the implected object is implected.

A *Sosein* function can first be defined exclusively for constitutive or nuclear properties in this way:

$$(\forall x)(\forall P_1)(\forall P_2)\dots(\forall P_n)\dots(Sosein(x) = \{P_1, P_2, \dots, P_n, \dots\} \\ (P_1x \& P_2x \,\&\, \dots P_nx \,\&\, \dots))$$

Implection (proper or improper) of an object in another object is formalized as:

$$(\forall x)(\forall y)(Implected!x, y \equiv Sosein(x) \in Sosein(y))$$

Proper and improper implection are distinguished by means of the following pairs of logically equivalent forms:

$$(\forall x)(\forall y)(Prop\text{-}implected!x, y \equiv (Implected!x, y \,\&\, -Implected!y, x))$$
$$(\forall x)(\forall y)(Improp\text{-}implected!x, y \equiv (Implected!x, y \,\&\, Implected!y, x))$$
$$(\forall x)(\forall y)(Prop\text{-}implected!x, y \equiv (Sosein(x) \in Sosein(y) \,\&\, Sosein(y) \notin Sosein(x)))$$
$$(\forall x)(\forall y)(Improp\text{-}implected!x, y \equiv (Sosein(x) \in Sosein(y) \,\&\, Sosein(y) \in Sosein(x)))$$

Where the *Soseine* of objects are as specified below, we can offer a perspicuous interpretation of the proper implection of o_1 in o_2 that, even in symbolic notation, is visually an entwining, interweaving, or embedding of the properties that constitute a properly implected object in the relatively more complete set of properties that constitute the object in which a properly implected object is properly implected. We then have:

$$Sosein(o_1) = \{P_1, P_2, \dots, P_i, \dots\}$$
$$Sosein(o_2) = \{P_1, P_2, \dots, P_i, \dots, P_n, \dots\}$$

We see that object o_1 is properly implected in object o_2 if and only if $Sosein(o_1) \in Sosein(o_2) \,\&\, Sosein(o_1) \notin Sosein(o_2)$. This provides an intuitive set

theoretical sense in which an incomplete properly implected object is analogically or metaphorically speaking embedded, interwoven, or entwined in a relatively more complete object, by the literal embedding, interweaving, or entwining of the members of the *Sosein* set of its constitutive properties in the *Sosein* set of constitutive properties of objects in which the properly implected object is properly implected. The same account explains implexive nonbeing as the embedding of the *Sosein* of an implected object in the *Sosein* of a beingless object. The implexive so-being of an implected object is the set of constitutive properties with which the implected object is associated by virtue of its implection in a particular object with those properties. The universal ball by this account has, not the so-being, but only the implexive so-being, of being red, if it happens to be properly implected in a particular red ball, and the implexive so-being of being small and blue, if it is implected in a particular small blue ball.[21]

Meinong's theory of implection now permits a more streamlined domain of objects than is usually countenanced even by the most nominalistic kinds of ontologies that admit abstract entities. In allowing incomplete and impossible nonexistent and nonsubsistent objects, first of all, contrary to a frequently voiced but entirely misinformed complaint, Meinong cannot be said to have planted a jungle, for he has not inflated ontology with explanatorily or otherwise theoretically unnecessary objects. The objects Meinong posits are needed to account for the intentionality or object-directedness of ordinary and scientific thought and discourse, which by his broadly Brentanian understanding of intentionality, cannot be explained except as about or directed toward beingless as well as existent and subsistent objects. Beingless objects do not inflate ontology, because they reside instead in the Meinongian semantic domain's extraontology. Meinong obviously does not claim that nonexistent nonsubsistent objects exist or subsist, or that beingless objects somehow after all have some sort of being. Nor can Meinong's object theory correctly be said to have planted a tangled jungle in the sense of admitting logically or metaphysically disorderly objects into the ontology or extraontology. Beingless Meinongian objects are identified and individuated, like existent and subsistent entities, by application of Leibniz's Law. Meinongian objects are identified and individuated on the basis of the constitutive properties truly predicated of them as constituting their unique being-independent being-indifferent *Soseine*. Meinong's object theory as a result is neither an excessively lush nor tangled ontic or semantic jungle.[22]

The distinction between Meinongian ontology and extraontology permits object theory to support an even more parsimonious ontology of existent or subsistent objects than Russell's later extensionalist theory of entities or 'individuals', including

[21] Meinong, *Über Möglichkeit und Wahrscheinlichkeit*, pp. 209-214, 715.

[22] The origin of the phrase 'Meinong's Jungle' seems to be William C. Kneale, *Probability and Induction* (Oxford: The Clarendon Press, 1949), p. 12. See Findlay, *Meinong's Theory of Objects and Values*, Preface, p. xi. Routley, *Exploring Meinong's Jungle and Beyond*, p. v. P.M.S. Hacker, *Insight and Illusion: Themes in the Philosophy of Wittgenstein*, revised edition (Oxford: The Clarendon Press, 1986), p. 8. Robert Campbell, "Did Meinong Plant a Jungle?", *Philosophical Papers*, 1, 1972, pp. 89-102. Jacquette, 'On Defoliating Meinong's Jungle', *Axiomathes*, special issue on 'The Philosophy of Alexius Meinong', 7, 1996, pp. 17–42.

universals, or Quine's extensionalist desert landscape ontology of spatiotemporal entities and abstract classes.[23]

The idea is to have available a nonexistent nonsubsistent abstract referent that does all the work of traditional existent or subsistent abstract entities, but for which ontic commitment is unnecessary. Meinong's extraontology makes it possible to go beyond concerns about whether Meinong planted a jungle to present an ontically more streamlined desert-like ontology of existent entities than that provided by even the most austere materially adequate extensionalist metaphysics.[24] Meinong himself, we should be clear, does not do this. He prefers to maintain as part of ontology many but not all traditional abstract entities. In a revisionary neo-Meinongian object theory, in contrast, we can adapt Meinong's theory in such a way that the ontology is redefined to contain all and only actual concrete existent entities, from which abstracta of all kinds are removed and relegated to a subdomain within the extraontology along with other incomplete objects and alongside impossible nonexistent nonsubsistent objects like the golden mountain and round square.[25]

Within a general Meinongian object theory framework, drawing on Meinong's concept of implexive being and nonbeing, it is possible to advance an absolutely minimalist Meinongian extraontology consisting of only one beingless object, which I shall call the *maximally impossible object*.[26] The concept of implexive being and nonbeing suggests an obvious strategy for extraontological reduction in Meinongian logic. The beingless objects of the extraontology can be reduced in number and made more orderly if at least some beingless objects are regarded as implected in others. We approach this sort of economy in the miniature by acknowledging the red round square as a beingless object in the extraontology, and interpreting references to the red round (object), round square (object), red square (object), round (object), square (object), and red (object), as implected in the red round square. The idea can then be taken to its logical extreme by introducing the con-

[23]Russell, 'The World of Universals', *The Problems of Philosophy* [1912] (New York: Oxford University Press, 1959), Chapter IX, pp. 95-97. Russell, "On the Relations of Universals and Particulars", *Proceedings of the Aristotelian Society*, 12, 1911-12, pp. 1-24. Quine, "On What There Is", *From a Logical Point of View* (Cambridge: Harvard University Press, 1953), pp. 12-15, 18. Quine, "Logic and the Reification of Universals", *From a Logical Point of View*, pp. 112-117.

[24]Jacquette, "Virtual Relations", *Idealistic Studies*, 25, 1995, pp. 141–154.

[25]I offer a proposal specifically for an ontically neutral object theory of relations in Jacquette, "Virtual Relations". The present suggestion generalizes the strategy to all abstracta. See my reply to William J. Rapaport in Jacquette and Rapaport, "Virtual Relations vs. Virtual Universals: Essay, Comments, and Reply" (with William J. Rapaport), Center for Cognitive Science Technical Report 95–10 (Buffalo: State University of New York Technical Report Series, 1995), pp. 1–20.

[26]I discuss the concept of a maximally impossible object in Jacquette, *Meinongian Logic*, pp. 103-104, 117, 125, 128, 135, 137, 139-140, 145. To be maximally impossible is not a constitutive property. But the phrase 'maximally impossible' is used in this context as a shorthand for a generalization of all constitutive properties and their complements, as in the red nonred, green nongreen, round nonround, square nonsquare, etc., object. See also, Jacquette, "On Defoliating Meinong's Jungle".

cept of a single beingless object of the Meinongian extraontology in which all other beingless objects are properly implected.

The maximally impossible object is the round square taken to its absolute predicational limit. The maximally impossible object is the beingless object that has in its *Sosein* every metaphysically possible and therefore collectively a maximally metaphysically impossible combination of constitutive properties. It is the nonexistent nonsubsistent object that is not only round and square, but golden and a mountain, nongolden and a nonmountain, fat, bald and a man, nonfat, nonbald, a nonman, and so on, for every property and property-complement pair. If the Meinongian extraontology admits the round square, it must surely admit the maximally impossible object, despite the fact that Meinong nowhere mentions such an object. Meinong's concept of implexive being and nonbeing implies that the maximally impossible object has implected within it every beingless possible and impossible object. It does so by embedding in its excessively rich *Sosein* the *Sosein* of every other beingless object. It is therefore possible for anyone concerned about the lushness of the revisionary Meinongian extraontology to impose an extraontological reduction on its contents, to reduce the sheer number of beingless possible and impossible objects it contains to just one, the maximally impossible object, in which every other beingless object is properly implected.[27]

From the maximally impossible object individual reference and true predication of constitutive properties to the beingless objects it implexively contains can be theoretically derived in much the same way that extensionalist semantic ontologists sometimes propose a reduction to and derivation of such entities as properties from sets or propositions, sets from propositions or properties, or properties from sets or propositions, and of some kinds of sets, properties, or propositions from other ontically more primitive kinds of sets, properties, or propositions. The *Sosein* of the round square, and of the golden mountain, the possible fat bald man in the doorway, the universal redness, and so on, and with them their distinguishing identity conditions, are all logically distinguishably properly implected in the *Sosein* of the maximally impossible object. Meinong's concept of implexive being and nonbeing can thus be used to reduce even the extraontology of beingless objects to the implexive nonbeing of possible beingless objects in existent (or subsistent, if these are allowed) objects, by virtue of the proper implexive so-being of all beingless objects in the so-being of the maximally impossible object, and of the improper implexive so-being of the maximally impossible object in itself. The maximally impossible object by this conception just is the coincidence in a single beingless object of every constitutive property and its complement. The Leibnizian identity conditions for all beingless objects, their uniquely individuating *Soseine*, can thus be defined in terms of selected subsets of constitutive properties and their complements implected in the *Sosein* of the maximally impossible object. The maximally impossible object literally contains within itself the implexive nonbeing of every beingless incomplete or impossible Meinongian object.

[27]See Jacquette, "Meinong's Concept of Implexive Being and Nonbeing", *Grazer Philosophische Studien*, 50, pp. 233–271.

Accordingly, we now add some new formalisms to the neo-Meinongian theory of implection. This makes it possible to express the implection of all beingless objects in a single maximally impossible object. We begin by regimenting some working assumptions about the concept of being, featuring equivalences involving the conditions of internal predicational completeness and consistency. As before, an exclamation mark attached to a predicate indicates an extranuclear or extraconstitutive property, and the absence of an exclamation mark indicates a constitutive or nuclear property.

$$(\forall x)(\mathit{Consistent}!x \equiv -(\exists P)(Px \,\&\, (\overline{P}x))$$
$$(\forall x)(\mathit{Consistent}!x \equiv -(\exists P)(P \in \mathit{Sosein}(x) \,\&\, (\overline{P} \in \mathit{Sosein}(x)))$$
$$(\forall x)(\mathit{Complete}!x \equiv (\forall P)(P \in \mathit{Sosein}(x) \vee (\overline{P} \in \mathit{Sosein}(x)))$$
$$(\forall x)(\mathit{Complete}!x \equiv (\forall P)(Px \vee (\overline{P}x))$$
$$(\forall x)(\mathit{Being}!x \equiv (\mathit{Complete}!x \,\&\, \mathit{Consistent}!x))$$
$$(\forall x)(\mathit{Being}!x \equiv (\forall P)((Px \vee (\overline{P}x) \,\&\, -(Px \,\&\, (\overline{P})x)))$$
$$(\forall x)(\mathit{Being}!x \equiv ((\forall P)(P \in \mathit{Sosein}(x) \vee (\overline{P} \in \mathit{Sosein}(x)) \,\&\, -(\exists P)(P \in \mathit{Sosein}(x) \,\&\, (\overline{P} \in \mathit{Sosein}(x))))$$
$$(\forall x)(\mathit{Being}!x \supset \mathit{Consistent}!x)$$
$$(\forall x)(\mathit{Being}!x \supset \mathit{Complete}!x)$$
$$-(\forall x)(\mathit{Complete}!x \supset \mathit{Being}!x)$$
$$-(\forall x)(\mathit{Consistent}!x \supset \mathit{Being}!x)$$

Now we are ready to define the concept of the maximally impossible object as the object that has in its *Sosein* every constitutive property and the complement of every constitutive property. We consider two alternative definitions that are logically equivalent by the concept of the *Sosein* function:

$$(\forall x)(\mathit{Maximpossible}!x \supset (\forall P)(Px \,\&\, (\overline{P}x))$$
$$(\forall x)(\mathit{Maximpossible}!x \supset (\forall P)(P \in \mathit{Sosein}(x) \& (\overline{P} \in \mathit{Sosein}(x)))$$

The *Sosein* function and definitions of the maximally impossible object, along with Leibnizian identity conditions for Meinongian objects, further imply that there is only one unique maximally impossible object:

$$(\forall x)(\forall y)((\mathit{Maximpossible}!x \,\&\, \mathit{Maximpossible}!y) \supset x = y)$$

We can therefore speak of *the* maximally impossible object. It is useful for this purpose to formalize the concept of a Meinongian theory of definite description:

$$(\forall F_1)...(\forall F_n)[F_1(\imath_m x(F_i x \,\&...\&\, F_n x) \equiv (\exists x)[(F_i x) \,\&...\&\, F_n x) \,\& \\ (\forall y)[(F_i y \,\&...\&\, F_n y) \equiv (x = y)] \,\&\, F_1 x]]$$

By comparison, classical Russellian definite description theory requires the extraconstitutive existence or being of the definitely described object to which any predicate truly applies:

$$(\forall F_1)...(\forall F_n)[F_1(\imath_r x(F_i x \&...\& F_n x)) \equiv (\exists x)[(\mathit{Being}!x \& F_i x) \&...\& F_n x) \& \\ (\forall y)[(F_i y \&...\& F_n y) \equiv (x = y)] \& F_1 x]]$$

This allows us to use the Meinongian definite descriptor '$\imath_m x$' by contrast with the Russellian '$\imath_r x$' to refer more succinctly to the maximally impossible object and formalize in several ways its logical singularity. Here we posit the maximally impossible object as belonging to the Meinongian domain by simple Meinongian quantification, and equivalently by Meinongian description and abstraction:

$$(\exists x)\mathit{Maximpossible}!x$$
$$(\exists x)(x = \imath_m y \mathit{Maximpossible}!y)$$
$$(\exists x)(\lambda y[y = \imath_m z \mathit{Maximpossible}!z]x)$$

The *Sosein* function and definition of implection, proper implection, and of the maximally impossible object now enables the logic to represent the implection of every beingless object in the maximally impossible object, and the proper implection of every incomplete object and every beingless object other than the maximally impossible object in the maximally impossible object. Here is a series of such formalizations:

$$(\forall x)(\mathit{Maximpossible}!x \equiv x = \imath_m y \mathit{Maximpossible}!y)$$
$$(\forall x)(\mathit{-Being}!x \supset \mathit{Implected}!x, \imath_m y \mathit{Maximpossible}!y)$$
$$(\forall x)(\mathit{-Complete}!x \supset \mathit{Prop\text{-}implected}!x, \imath_m y \mathit{Maximp}!y)$$
$$(\forall x)((\mathit{-Being}!x \,\&\, -\mathit{Maximpossible}!x) \supset \mathit{Implected}!x, \imath_m\ y \mathit{Maximpossible}!y)$$
$$-(\exists x)(\mathit{-Being}!x \,\&\, -\mathit{Implected}!x, \imath_m y \mathit{Maximpossible}!y)$$
$$-(\exists x)(-\mathit{Comp}!x \,\&\, -\mathit{Prop\text{-}implected}!x, \imath_m y \mathit{Maximpossible}!y)$$
$$-(\exists x)(-\mathit{Being}!x \,\&\, -\mathit{Maximpossible}!x \,\&\, -\mathit{Implected}!x, \imath_m y \mathit{Maximp}!y)$$
$$-(\exists x)(\mathit{Being}!x \,\&\, \mathit{Implected}!x, \imath_m y \mathit{Maximp}!y)$$
$$-(\exists x)(\forall y)(\mathit{Being}!x \,\&\, \mathit{Prop\text{-}implected}!x, y)$$

The latter expressions by the definition of implection and proper implection imply that there is no singularity of beinglessness, but a plurality of beingless objects in neo-Meinongian extraontology. The simplest proof is to consider that by free assumption we have beingless objects that by Leibnizian identity conditions are distinct from or other than the maximally impossible object, so that not every beingless object is strictly identical with, despite being implected in, the maximally impossible object. Again, we have a variety of ways to express these conclusions:

$$(\exists x)(\mathit{-Being}!x \,\&\, \mathit{Prop\text{-}implected}!x, \imath_m y \mathit{Maximpossible}!y)$$
$$(\exists x)(\mathit{-Being}!x \,\&\, -\mathit{Prop\text{-}implected}!, \imath_m y \mathit{Maximpossible}!y, x)$$
$$(\exists x)(\mathit{-Being}!x \,\&\, -\mathit{Maximpossible}!x)$$
$$-(\forall x)(\mathit{-Being}!x \supset \mathit{Maximpossible}!x)$$
$$(\exists x)(\forall y)(\mathit{-Being}!x \,\&\, \mathit{Maximpossible}!y \,\&\, x \neq y)$$
$$-(\exists x)(\lambda y[y = \imath_m z - \mathit{Being}!z]x)$$
$$-(\forall x)(\mathit{-Being}!x \equiv x = \imath_m y - \mathit{Being}!y)$$
$$(\exists x)(\mathit{Pimp}!x, \imath_m y \mathit{Maximpossible}!y \,\&\, -\mathit{Being}!x)$$
$$(\exists x)(\mathit{-Being}!x \,\&\, x \neq \imath_m y \mathit{Maximpossible}!y)$$
$$(\exists x)(\exists y)(\mathit{-Being}!x \,\&\, -\mathit{Being}!y \,\&\, x \neq y)$$

The greater theoretical economy of a neo-Meinongian logic and semantics should now be evident. There are just as many beingless objects in the new system as in Meinong's original extraontological jungle, for we have said that implection and implexive being like implexive nonbeing should not be understood as a kind or mode of being, but as a kind or mode of nonbeing.

Extraontological economy is achieved in a neo-Meinongian semantic domain if reference to beingless objects in (the equivalent of the) axioms of the new object theory or reductive metatheory are reduced to just one, the maximally impossible object. We know that there most definitely is only one such beingless object in the object theory domain, and that all other beingless objects are properly implected in it — including all incomplete or impossible, nonexistent or nonsubsistent Meinongian objects. The beinglessness of all beingless objects implected in the maximally impossible object is confirmed by the set theoretical interpretation of the entwining or embedding of the elements of the *Sosein* set of the constitutive properties of the implected object among the elements of the *Sosein* set of the constitutive properties of the object in which the implected object is implected.

We can derive every beingless object from the singular maximally impossible object in which all are implected. In this sense, as in other ontic reductions in non-Meinongian extensional ontologies, a neo-Meinongian object theory or extraontological metatheory need not acknowledge more than one beingless object in its primitive terms and minimal ontic and extraontic domain commitments as belonging to its extraontology. All other beingless objects are properly implected in it, from which they can all therefore be theoretically derived. An economy results even in an extensional ontology by adopting a version of Aristotle's inherence theory to avoid explicit ontic commitment to the existence or subsistence of universals, from which they can then be abstracted, or by other ontically reductive strategies. In much the same way, a neo-Meinongian object theory or intensional ontology and extraontology accomplishes an economizing of the extraontology by adopting a version of Meinong's implection theory (which is also though it is somewhat more than an inherence and inhesion theory) and the concepts of implexive being and nonbeing as two kinds or modes of nonbeing, allowing all beingless objects to be implected in the singular maximally impossible object from which they can similarly be derived.

It must be emphasized that Meinong ventured no such application of his theory of implexive being, so-being, and nonbeing. The concepts as he presents them nevertheless seem naturally to lend themselves to this kind of reduction. We can think of the Meinongian semantic domain as divided into two main parts. The ontology consists only of actually existent spatiotemporal entities, a truly desert landscape. The extraontology reductively speaking need consist only of the beingless maximally impossible object, within the so-being of which the so-being of every other beingless object is properly implected. Here, then, is surely no ontological or extraontological lush or tangled jungle. The neo-Meinongian ontology, consisting only of spatiotemporal existents, is more austere and desert-like than even the most economical materially adequate extensional semantic domain. The

neo-Meinongian extraontology is well-ordered by the identity conditions that associate via the domain comprehension principle a unique beingless object with every metaphysically incomplete or impossible set of constitutive properties. Rather every beingless object is identified and individuated by virtue of its unique combination of constitutive properties or their complements, and every such combination is to be found implected within and can therefore be derived from the so-being of the maximally impossible beingless object, the only object that extraontologically primitively or reductively need be considered as belonging to the most austere imaginable neo-Meinongian extraontology.

What, however, of the proper implection of existent and subsistent objects in the maximally impossible object? Meinong does not consider this application of the concept of implection, but it arises on the set theoretical interpretation even for complete but impossible Meinongian objects other than the maximally impossible object. An illustration is the proper implection of Meinong's friend's billiard ball in the beingless impossible Meinongian object whose *Sosein* includes all of the constitutive properties in the *Sosein* of Meinong's friend's billiard ball plus the complement of at least one of its constitutive properties. There are two directions in which this implication can be pursued. Meinong does not seriously entertain the possibility that existent or subsistent objects might be properly implected in beingless objects. This oversight might be explained in part by Meinong's general lack of interest in predicationally over-determined or complete but impossible objects. If we choose to preserve Meinong's restriction of implective being and nonbeing to the implection of beingless objects only in existent, subsistent, or predicationally relatively more complete but still beingless objects, then we can redefine (proper or improper) implection in this way:

$$(\forall x)(\forall y)(\mathit{Implected}!x, y \supset (\mathit{–Being}!x \,\&\, \mathit{Sosein}(x) \in \mathit{Sosein}(y)))$$

The specific definitions of proper and improper implection can remain just as they are formulated above, with correspondingly restricted entailments determined by the limitation of implection to beingless Meinongian objects. This restricted form of implection reflects a relatively metaphysically robust sense of implexive being, by which implection of objects with being in any other object is forbidden, and as such is more closely akin to Aristotelian inherence. It might be further justified by the need to distinguish between the implection of objects in other objects and the mereological physical containment of existent parts in existent wholes.[28]

[28]Peter M. Simons, "Free Part-Whole Theory", in Karel Lambert, ed., *Philosophical Applications of Free Logic* (Oxford: Oxford University Press, 1991), pp. 285-305. See section 5 on 'Implexive Containment', pp. 294-296. Simons argues that Meinongian implection of incomplete in existents or subsistents has all the formal properties of mereological containment, provided that the nuclear-extranuclear property distinction is enforced, but is not itself a genuine mereological part-whole relation. Simons' example involves the existence of the proper parts of George Washington and the nonexistence but only implexive existence of George Washington's implexive parts. The problem is clearly related to Meinong's discussion of the relative degrees of possibility

If it is judged philosophically more correct to recognize implection as obtaining wherever the *Sosein* set of constitutive properties of an object regardless of its ontic status is contained in the *Sosein* set of an object regardless of its ontic status, then we can preserve the original unrevised definition of implection. To do so is to permit the proper implection of existent and subsistent objects in beingless overdetermined or predicationally complete but impossible Meinongian objects, including the maximally impossible object. This ontically unrestricted form of Meinongian implection is not nearly as metaphysically robust as the ontically restricted concept in its similarity to Aristotelian inherence. It leads directly to a Meinongian version of Parmenideanism, by which the phenomenal world of physical or material spatiotemporal objects is implected in the maximally impossible object as an all-absorbing One. Here we have implection as a purely abstract logical or set theoretical relation, without ontic implications, in which the existent and subsistent are properly implected in certain beingless objects and reductively in the One maximally impossible object.[29]

Parmenides, as the fragments suggest, also believed that the phenomenal world is nonexistent because of its ephemeral ever-changing fluxlike nature.[30] This idealist conception can also be accommodated in the ontically unrestricted formulation of Meinong's concept of implexive being and nonbeing. We might appeal for such a conclusion to the essential predicational incompleteness of physical objects implied by the indeterminacy thesis in quantum physics. If spatiotemporal objects at the microphysical level consist of quanta that necessarily lack either determinate position or determinate momentum or velocity, then all of what we are accustomed to think of as existent physical reality by Meinong's theory of beingless incomplete objects is actually nonexistent.[31] If 'existent' spatiotemporal or physical 'reality', consisting of the entities we regard as constituting the furniture of the universe, is

---

of instantiation of a Goethe or a Beethoven. Here we must distinguish between Goethe and Beethoven on the one hand, and whatever subset of the *Sosein* set of constitutive properties of these persons might be thought sufficient to instantiate an instance of a Goethe or a Beethoven. It is not immediately clear how this might be done, except by nominal stipulation of a set of essential properties that fall short of the complete *haecceity* of Goethe himself or Beethoven himself. For other incomplete but not impossible objects like the golden mountain or Pegasus, there is no clearcut basis by which to regard such objects as something more individual and less universal than such standard examples of universals as the ideal state or the color blue.

[29]The beingless maximally impossible object as the Meinongian One is in other ways obviously quite different from the Parmenidean concept of the One interpreted as the only Reality.

[30]Parmenides' fragments are collected in *Ancilla to the Pre-Socratic Philosophers: A Complete Translation of the Fragments in Diels, Fragmente der Vorsokratiker*, Kathleen Freeman, ed. and trans. (Cambridge: Harvard University Press, 1957), pp. 41-51.

[31]Werner Heisenberg, *The Physical Principles of the Quantum Theory*, Carl Eckart and Frank C. Hugh, trans. (New York: Dover Publications, Inc., 1930). Heisenberg, *Physics and Philosophy: The Resolution in Modern Science* (London: Allen and Unwin, 1959). Peter Forrest, *Quantum Metaphysics* (Oxford: Basil Blackwell, 1988), especially pp. 25-45, 102-122. An elegant formal description implying the predicational incompleteness of quantum phenomena in terms of probability t-cones is given by Bas C. van Fraassen, *Quantum Mechanics: An Empiricist View* (Oxford: The Clarendon Press, 1991), pp. 51-53. See Jagdish Mehra, *The Quantum Principle: Its Interpretation and Epistemology* (Dordrecht: D. Reidel Publishing Company, 1974), pp. 107-116.

predicationally incomplete and hence nonexistent, as Parmenides for different but related reasons also taught, then all beingless objects of the phenomenal world are properly implected in a single beingless object, the maximally impossible object, as the Meinongian-Parmenidean One.

The philosophical success of Meinong's object theory is seen in its fundamental ontic neutrality as the semantics of all possible thought and expression. Providing the semantics for such opposite metaphysics of the phenomenal world as those of Parmenides and Aristotle is the most demanding test of the ontic neutrality of Meinong's object theory. Through its concept of implexive being and nonbeing, Meinong's theory makes possible by means of special provision among many other items in a spectrum of possibilities both the formulation of a Parmenidean notion of the nonexistence of the phenomenal world and the collection of all that is thinkable into an all-embracing One, and of the Aristotelian thesis of the phenomenal world as the only reality consisting of primary substances or discrete existent physical entities. We should expect Meinong's object theory to provide the semantics for these among many other correct or incorrect metaphysical systems, and we should be prepared to criticize it if it did not.

Still, would it not be too remarkable if Meinong's Brentanian Aristotelianism as a methodological superstructure built on ontically neutral object theory foundations, taken to its logical extreme, led back to Parmenides by way of the empiricist epistemic limitations embodied in the indeterminacy principle of modern quantum physics to a metaphysics of the (beingless) One and the world of empirical phenomena as nonexistent because fundamentally predicationally incomplete? Meinong avoids such inferences in the metaphysics he superimposes on his ontically neutral object theory only by refusing to recognize the implection of any existent or subsistent objects in other existent, subsistent, or beingless objects. We can express Meinong's implicit stipulation formally as proposed by means of a restriction on the *Sosein* set theoretical interpretation of implection, allowing only beingless objects to be implected in other objects. However mad or intellectually objectional we may find Parmenidean idealism as opposed to Aristotelian realism to be, it is a sign of the ontic neutrality of Meinong's object theory semantics as the most general theory of meaning that it provides an absolutely ontically neutral foundation for the most diverse conceptions of metaphysics by means of its categories of the extraontology and concept of implexive being and nonbeing.

ACKNOWLEDGEMENTS

I am grateful to the Alexander von Humboldt-Stiftung for supporting my research on Meinong and intensional logics in Germany in 1989-1990 at the Universität Mannheim and 2000–2001 at the at the Franz Brentano Forschung, Bayerische-Julius-Maximilians-Universität, Würzburg. This essay was completed during my Resident Research Fellowship at the Netherlands Institute for Advanced Study (NIAS), Royal Netherlands Academy of the Arts and Sciences (KNAW), 2005–2006. Ancestors of the formalisms presented here have appeared in my *Meinongian*

Logic, "Object Theory Foundations for Intensional Logic", and "Meinong's Concept of Implexive Being and Nonbeing"; see References below and in the Notes. I thank the series editors, Dov Gabbay and John Woods, for their kind invitation to contribute an essay on the logic of Meinong's object theory semantics, and Jane Spurr at the Department of Computer Science, King's College London for patient reworking of a challenging electronic manuscript.

BIBLIOGRAPHY

[Abertazzi *et al.*, 1996] L. Albertazzi, M. Libardi, and R. Poli, eds. *The School of Franz Brentano*, Dordrecht and Boston, Kluwer Academic Publishers, 1996.

[Barber, 1967] K. F. Barber. *Meinong's 'Hume Studies': Translation and Commentary*, 1966 Ph.D. thesis, The University of Iowa. Ann Arbor: University Microfilms, Inc, 1967.

[Barber, 1970] K. F. Barber. Meinong's Hume Studies, Part I: Meinong's Nominalism; Part II: Meinong's Analysis of Relations, *Philosophy and Phenomenological Research*, 30, 550-567; 31, 1971, 564-584, 1970.

[Berkeley, 1949] G. Berkeley. *The Works of George Berkeley Bishop of Cloyne*, edited by A.A. Luce and T.E. Jessop. London: Thomas Nelson & Sons. *A Treatise on the Principles of Human Knowledge*; *Three Dialogues Between Hylas and Philonous*, *Works*, Vol. II, 1949.

[Campbell, 1972] R. Campbell. Did Meinong Plant a Jungle? *Philosophical Papers*, 1, 89-102, 1972.

[Carnap, 1956] R. Carnap. *Meaning and Necessity*: *A Study in Semantics and Modal Logic*, enlarged edition, Chicago, University of Chicago Press, 1956.

[Chisholm, 1972] R. M. Chisholm. Beyond Being and Non-Being, *Philosophical Studies*, 24, 245-255, 1972.

[Chisholm, 1982a] R. M. Chisholm. *Brentano and Meinong Studies*, Atlantic Highlands, Humanities Press, 1982.

[Chisholm, 1982b] R. M. Chisholm. Converse Intentional Properties, *The Journal of Philosophy*, 79, 537-545, 1982.

[Findlay, 1995] J. N. Findlay. *Meinong's Theory of Objects and Values*, *Edited with an Introduction by Dale Jacquette*, from the second edition, Oxford University Press, 1963, Aldershot, Ashgate Publishing (Gregg Revivals), 1995.

[Fine, 1984] K. Fine. Critical Review of Parsons' *Nonexistent Objects*, *Philosophical Studies*, 45, 95-142, 1984.

[Forrest, 1988] P. Forrest. *Quantum Metaphysics*. Oxford: Basil Blackwell, 1988.

[Freeman, 1957] K. Freeman, editor and translator. *Ancilla to the Pre-Socratic Philosophers: A Complete Translation of the Fragments in Diels, Fragmente der Vorsokratiker*. Cambridge: Harvard University Press, 1957.

[Griffin, 1979] N. Griffin. The Independence of *Sosein* from *Sein*, *Grazer Philosophische Studien*, 9, 23-34, 1979.

[Grossmann, 1974a] R. Grossmann. Meinong's Doctrine of the *Aussersein* of the Pure Object, *Noûs*, 8, 67-82, 1974.

[Grossmann, 1974b] R. Grossmann. *Meinong*, London, Routledge & Kegan Paul, 1974.

[Hacker, 1986] P. M. S. Hacker. *Insight and Illusion: Themes in the Philosophy of Wittgenstein*, revised edition. Oxford: The Clarendon Press, 1986.

[Haller, 1972] R. Haller, ed. *Jenseits von Sein und Nichtsein*, Graz, Akademische Druck- und Verlagsanstalt, 1972.

[Heisenberg, 1930] W. Heisenberg. *The Physical Principles of the Quantum Theory*, Carl Eckart and Frank C. Hugh, trans. New York: Dover Publications, Inc, 1930.

[Heisenberg, 1959] W. Heisenberg. *Physics and Philosophy: The Resolution in Modern Science*. London: Allen and Unwin, 1959.

[Henkin, 1949] L. Henkin. The Completeness of the First–Order Functional Calculus, *Journal of Symbolic Logic*, 14, 159–166, 1949.

[Hume, 1975] D. Hume. *An Enquiry Concerning Human Understanding*, in *David Hume: Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, edited by L.A. Selby-Bigge, third edition, revised by P.H. Nidditch. Oxford: The Clarendon Press, 1975.

[Hume, 1978] D. Hume. *A Treatise of Human Nature*, edited by Selby-Bigge, second edition, revised by P.H. Nidditch. Oxford: The Clarendon Press, 1978.

[Jacquette, 1982] D. Jacquette. Meinong's Theory of Defective Objects, *Grazer Philosophische Studien*, 15, 1-19, 1982.

[Jacquette, 1986] D. Jacquette. Meinong's Doctrine of the Modal Moment, *Grazer Philosophische Studien*, 25-6, 423-438, 1986.

[Jacquette, 1987] D. Jacquette. Twardowski on Content and Object, *Conceptus: Zeitschrift für Philosophie*, 21, 193-199, 1987.

[Jacquette, 1988] D. Jacquette. Review of Karel Lambert, *Meinong and the Principle of Independence: Its Place in Meinong's Theory of Objects and its Significance in Contemporary Philosophical Logic*, *International Studies in Philosophy*, 20, 92–93, 1988.

[Jacquette, 1989a] D. Jacquette. Modal Meinongian Logic, *Logique et Analyse*, 125-26, 113-130, 1989.

[Jacquette, 1989b] D. Jacquette. Mally's Heresy and the Logic of Meinong's Object Theory, *History and Philosophy of Logic*, 10, 1-14, 1989.

[Jacquette, 1989c] D. Jacquette. Intentional Semantics and the Logic of Fiction, *The British Journal of Aesthetics*, 29, 168–176, 1989.

[Jacquette, 1989d] D. Jacquette, trans. Ernst Mally, 'On the Objects' Independence from Thought' ('Über die Unabhängigkeit der Gegenstände vom Denken'), *Man and World*, 22, 215-231, 1989.

[Jacquette, 1990] D. Jacquette. Review of Marie–Luise Schubert Kalsi, *Meinong's Theory of Knowledge*, *Noûs*, 24, 487–492, 1990.

[Jacquette, 1990–91] D. Jacquette. The Origins of *Gegenstandstheorie*: Immanent and Transcendent Intentional Objects in Brentano, Twardowski, and Meinong, *Brentano Studien*, 3, 277-302, 1990–91.

[Jacquette, 1991a] D. Jacquette. Mally, Ernst, in *Handbook of Metaphysics and Ontology*, Hans Burkhardt and Barry Smith, ed., Munich and Vienna, Philosophia Verlag, 485-486, 1991.

[Jacquette, 1991b] D. Jacquette. Review of Edward N. Zalta, *Intensional Logic and the Metaphysics of Intentionality*, *Philosophy and Phenomenological Research*, 51, 439–444, 1991.

[Jacquette, 1992] D. Jacquette. Meinongian Models of Scientific Law, in Paśniczek, ed., 86-104, 1992.

[Jacquette, 1994a] D. Jacquette. *Philosophy of Mind*, Englewood Cliffs, Prentice Hall, 1994.

[Jacquette, 1994b] D. Jacquette. Meinongian Logic and Anselm's Ontological Proof for the Existence of God, *The Philosophical Forum*, 25, 231-240, 1994.

[Jacquette, 1994c] D. Jacquette. Tarski's Quantificational Semantics and Meinongian Object Theory Domains, 75, 88-107, 1994.

[Jacquette, 1994d] D. Jacquette. A Meinongian Theory of Definite Description, *Axiomathes*, 5, 345-359, 1994.

[Jacquette, 1995a] D. Jacquette. Brentano, Franz, Meinong, Alexius, in *The Cambridge Dictionary of Philosophy*, Robert Audi, ed., Cambridge, Cambridge University Press, 1995.

[Jacquette, 1995b] D. Jacquette. Findlay and Meinong, in Findlay, *Meinong's Theory of Objects and Values, With an Introduction by Dale Jacquette*, 1995.

[Jacquette, 1995c] D. Jacquette. Meinong's Concept of Implexive Being and Nonbeing, *Grazer Philosophische Studien*, 50, 233–271, 1995.

[Jacquette, 1995d] D. Jacquette. Object Theory Foundations for Intensional Logic, *Acta Analytica*, 13, 33–63, 1995.

[Jacquette, 1995e] D. Jacquette. Virtual Relations, *Idealistic Studies*, 25, 141–154, 1995.

[Jacquette, 1996a] D. Jacquette. *Meinongian Logic: The Semantics of Existence and Nonexistence*, Berlin and New York, Walter de Gruyter, 1996.

[Jacquette, 1996b] D. Jacquette. On Defoliating Meinong's Jungle, *Axiomathes*, 7, 17–42. 1996.

[Jacquette, 1996c] D. Jacquette. Alexius Meinong (1853-1920), in *The School of Franz Brentano*, edited by Albertazzi, Libardi, and Poli, 131–159, 1996.

[Jacquette, 1997a] D. Jacquette. Conceivability, Intensionality, and the Logic of Anselm's Modal Argument for the Existence of God, *International Journal for Philosophy of Religion*, 42, 163–173, 1997.

[Jacquette, 1997b] D. Jacquette. Reflections on Mally's Heresy, *Axiomathes*, 8, 163–180, 1997.
[Jacquette, 1999a] D. Jacquette. Review of Marie-Luise Schubert Kalsi, *Alexius Meinong's Elements of Ethics, With a Translation of the Fragment Ethische Bausteine*, *Review of Metaphysics*, 52, 727–730, 1999.
[Jacquette, 1999b] D. Jacquette. Review of Jacek Paśniczek, *The Logic of Intentional Objects: A Meinongian Version of Classical Logic*, *Journal of Symbolic Logic*, 64, 1847–1849, 1999.
[Jacquette, 2000a] D. Jacquette. Confessions of a Meinongian Logician, *Grazer Philosophische Studien*, 58–59, 151–180, 2000.
[Jacquette, 2000b] D. Jacquette. An Internal Determinacy Metatheorem for Lukasiewicz's *Aussagenkalküls*, *Bulletin of the Section of Logic*, 29, 115–124, 2000.
[Jacquette, 2001a] D. Jacquette. *The School of Alexius Meinong*, edited with Liliana Albertazzi and Roberto Poli, Aldershot, Ashgate Publishing, 2001.
[Jacquette, 2001b] D. Jacquette. 'Truth and Fiction in David Lewis's Critique of Meinongian Semantics, *Metaphysica: International Journal for Ontology and Metaphysics*, 2, 73–106, 2001.
[Jacquette, 2003a] D. Jacquette. David Lewis on Meinongian Logic of Fiction, *Writing the Austrian Traditions: Relations Between Philosophy and Literature*, edited by Wolfgang Huemer and Marc-Oliver Schuster. Toronto: University of Toronto Press,Wirth-Institute for Austrian and Central European Studies, University of Alberta, 101–119, 2003.
[Jacquette, 2003b] D. Jacquette. Meinong on the Phenomenology of Assumption, *Studia Phaenomenologica*, 3, 155-177, 2003.
[Jacquette, 2006] D. Jacquette. Twardowski, Brentano's Dilemma, and the Content-Object Distinction, *Actions, Products and Things: Brentano and Polish Philkosophy*, Mind and Phenomenology series, edited by Arkadiusz Chrudzimski and Dariusz Lukasiewicz. Frankfurt: Ontos Verlag, 2006, pp 9–33.
[Kant, 1965] I. Kant. *Critique of Pure Reason*, trans. Norman Kemp Smith, New York, St. Martin's Press, 1965.
[Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics*. Amsterdam: North–Holland Publishing Co, 1952.
[Kneale, 1949] W. C. Kneale. *Probability and Induction*. Oxford: The Clarendon Press, 1949.
[Lambert, 1982] K. Lambert. A Logical Interpretation of Meinong's Principle of Independence, *Topoi*, 1, 87-96, 1982.
[Lambert, 1983] K. Lambert. *Meinong and the Principle of Independence: Its Place in Meinong's Theory of Objects and its Significance in Contemporary Philosophical Logic*, Cambridge, Cambridge University Press, 1983.
[Leblanc *et al.*, 1982] H. Leblanc, H. Goldberg, and G. Weaver. A Strong Completeness Theorem for Three–Valued Logic: Part I, in Leblanc, 1982b, 240–246, 1982.
[Leblanc, 1982a] H. Leblanc. A Strong Completeness Theorem for Three–Valued Logic: Part II, in Leblanc, 1982b, 247–257, 1982.
[Leblanc, 1982b] J. Leblanc. *Existence, Truth, and Provability*. Albany: State University of New York Press, 1982.
[Łukasiewicz, 1930] J. Łukasiewicz. Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls, *Comptes rendus des seánces de la Société des sciences et des lettres de Varsovie*, Classe III, 23, 51–77, 1930.
[Mally, 1904] E. Mally. Untersuchungen zur Gegenstandstheorie des Messens, in Meinong, ed., *Untersuchungen zur Gegenstandstheorie und Psychologie*. Leipzig: Verlag von Johann Ambrosius Barth, 1904.
[Mally, 1912] E. Mally. *Gegenstandstheoretische Grundlagen der Logik und Logistik*, Leipzig, Verlag von Johann Ambrosius Barth, 1912.
[McCall, 1967] S. McCall, ed. *Polish Logic: 1920–1939*. Oxford: Oxford University Press, 1967.
[Mehra, 1974] J. Mehra. *The Quantum Principle: Its Interpretation and Epistemology*. Dordrecht: D. Reidel Publishing Company, 1974.
[Meinong, 1921] A. Meinong. Zur Gegenstandstheorie, in Raymund Schmidt, ed., *Die Philosophie der Gegenwart in Selbstdarstellungen*. Leipzig: Felix Meiner Verlag, 1921.
[Meinong, 1965] A. Meinong. *Philosophenbriefe: aus der wissenschaftlichen Korrespondenz*, Rudolf Kindinger, ed., Graz, Akademische Druck- u. Verlagsanstalt, 1965.
[Meinong, 1969–78] A. Meinong. *Alexius Meinong Gesamtausgabe*, I-VIII, Rudolf Haller and Rudolf Kindinger in collaboration with Roderick M. Chisholm, ed., Graz, Akademische Druck- u. Verlagsanstalt, 1969–1978.

[Parsons, 1965] C. D. Parsons. Frege's Theory of Number in M. Black, ed., *Philosophy in America*. Ithaca: Cornell University Press, 1965.
[Parsons, 1974] T. Parsons. A Prolegomenon to Meinongian Semantics, *The Journal of Philosophy*, 71, 561-580, 1974.
[Parsons, 1975] T. Parsons. A Meinongian Analysis of Fictional Objects, *Grazer Philosophische Studien*, 1, 73-86, 1975.
[Parsons, 1978] T. Parsons. Nuclear and Extranuclear Properties, Meinong, and Leibniz, *Noûs*, 12, 137-151, 1978.
[Parsons, 1980] T. Parsons. *Nonexistent Objects*, New Haven, Yale University Press, 1980.
[Paźniczek, 1992] J. Paśniczek, ed. *Theories of Objects: Meinong and Twardowski*, Lublin, Wydawnictwo Uniwersytetu Marii Curie–Sklodowskiej, 1992.
[Paśniczek, 1998] J. Paśniczek. *The Logic of Intentional Objects: A Meinongian Version of Classical Logic*, Dordrecht and Boston, Synthese Library, Kluwer Academic Publishing, 1998.
[Perszyk, 1993] K. J. Perszyk. *Nonexistent Objects: Meinong and Contemporary Philosophy*, Dordrecht, Martinus Nijhoff, 1993.
[Quine, 1953a] W. V. O. Quine. On What There Is, *From a Logical Point of View*. Cambridge: Harvard University Press, 1953.
[Quine, 1953b] W. V. O. Quine. Logic and the Reification of Universals, *From a Logical Point of View*. Cambridge: Harvard University Press, 1953.
[Quine, 1960] W. V. O. Quine. *Word and Object*. Cambridge: The MIT Press, 1960.
[Quine, 1969] W. V. O. Quine. *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969.
[Rapaport, 1978] W. J. Rapaport. Meinongian Theories and a Russellian Paradox, *Noûs*, 12, 153-180, 1978.
[Reicher, 1998] M. E. Reicher. Die Logik des Nichtseienden, *Grazer Philosophische Studien*, 54, 165-196, 1998.
[Rescher, 1969] N. Rescher. *Many–Valued Logic*. New York: McGraw–Hill, 1969.
[Routley, 1981] R. Routley. *Exploring Meinong's Jungle and Beyond*, interim edition, Canberra, Australian National University, 1981.
[Russell, 1904] B. Russell. Meinong's Theory of Complexes and Assumptions, *Mind*, 13, 204-219, 1904.
[Russell, 1905a] B. Russell. On Denoting, *Mind*, 14, 479-493, 1905.
[Russell, 1905b] B. Russell. Review of A. Meinong, *Untersuchungen zur Gegenstandstheorie und Psychologie*, *Mind*, 14, 530-538, 1905.
[Russell, 1907] B. Russell. Review of A. Meinong, *Über die Stellung der Gegenstandstheorie im System der Wissenschaften*, *Mind*, 16, 436-439, 1907.
[Russell, 1911–12] B. Russell. On the Relations of Universals and Particulars, *Proceedings of the Aristotelian Society*, 12, 1-24, 1911–1912.
[Russell, 1912] B. Russell. The World of Universals, *The Problems of Philosophy*. New York: Oxford University Press, 1912.
[Russell, 1919] B. Russell. *Introduction to Mathematical Philosophy*. New York: Simon and Schuster, 1919.
[Ryle, 1933] G. Ryle. Review of J.N. Findlay, *Meinong's Theory of Objects*, *Oxford Magazine*, 52, 118-120, 1933.
[Ryle, 1972] G. Ryle. Intentionality-Theory and the Nature of Thinking, in Haller, ed., *Jenseits von Sein und Nichtsein*, 7-14, 1972.
[Shoenfield, 1967] J. R. Shoenfield. *Mathematical Logic*. Reading: Addison–Wesley Publishing Company, 1967.
[van Fraassen, 1991] B. C. van Fraassen. *Quantum Mechanics: An Empiricist View*. Oxford: The Clarendon Press, 1991.

THE LOGIC OF BROUWER AND HEYTING

Joan Rand Moschovakis

Intuitionistic logic consists of the principles of reasoning which were used informally by L. E. J. Brouwer, formalized by A. Heyting (also partially by V. Glivenko), interpreted by A. Kolmogorov, and studied by G. Gentzen and K. Gödel during the first third of the twentieth century. Formally, intuitionistic first-order predicate logic is a proper subsystem of classical logic, obtained by replacing the law of excluded middle $A \vee \neg A$ by $\neg A \supset (A \supset B)$; it has infinitely many distinct consistent axiomatic extensions, each necessarily contained in classical logic (which is axiomatically complete). However, intuitionistic second-order logic is inconsistent with classical second-order logic. In terms of expressibility, the intuitionistic logic and language are richer than the classical; as Gödel and Gentzen showed, classical first-order arithmetic can be faithfully translated into the negative fragment of intuitionistic arithmetic, establishing proof-theoretical equivalence and clarifying the distinction between the classical and constructive consequences of mathematical axioms.

The logic of Brouwer and Heyting is effective. The conclusion of an intuitionistic derivation holds with the same degree of constructivity as the premises. Any proof of a disjunction of two statements can be effectively transformed into a proof of one of the disjuncts, while any proof of an existential statement contains an effective prescription for finding a witness. The negation of a statement is interpreted as asserting that the statement is not merely false but absurd, i.e., leads to a contradiction. Brouwer objected to the general law of excluded middle as claiming *a priori* that every mathematical problem has a solution, and to the general law $\neg\neg A \supset A$ of double negation as asserting that every consistent mathematical statement holds.

Brouwer called "the First Act of Intuitionism" the separation of mathematics (an exclusively mental activity) from language. For him, all mathematical objects (including proofs) were mental constructions carried out by individual mathematicians. This insight together with "the intuition of the bare two-oneness" guided the creation of intuitionistic logic and arithmetic, with full induction and a theory of the rational numbers. It was limited, by Brouwer's early insistence on predicative definability, to the construction of entities which were finitely specifiable and thus at most denumerably infinite in number; this seemed to exclude the possibility of a good intuitionistic theory of the continuum.

A later insight led Brouwer to "the Second Act of Intuitionism," which accepted infinitely proceeding sequences of independent choices of objects already

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

constructed (*choice sequences* of natural numbers, for example) as legitimate mathematical entities. He succeeded in extracting positive information about the structure of the continuum from the fact that an individual choice sequence may be known only by its finite approximations. His "Bar Theorem" (actually a new mathematical axiom) essentially accepted induction up to any denumerable ordinal, significantly extending the intuitionistically available part of classical arithmetic and analysis. His Continuity Principle, which contradicted classical logic (but not classical first-order arithmetic), made possible the development of a rich alternative theory of the continuum. Brouwer's constructive understanding of the universal and existential quantifiers led him to accept some choice principles stronger than those of classical mathematics, and at the same time to reject the full classical axiom of choice.

Because intuitionistic logic is an *ex post facto* construction reflecting the forms of reasoning actually used in intuitionistic mathematics, in order to give a full account of the logic we will need to consider some of the mathematics as well. And because "the possibilities of thinking cannot be reduced to a finite number of rules constructed in advance,"[1] we would do well to remember that this is not the last word in the subject.

1 HISTORICAL DEVELOPMENT OF INTUITIONISTIC LOGIC

The Dutch mathematician and logician Luitzen Egbertus Jan Brouwer was born February 27, 1881 in Overschie, near Rotterdam, and died December 2, 1966 in Blaricum, east of Amsterdam. He was a highly original thinker who developed an intuitionistic alternative to classical mathematics, introduced the term "metamathematics" for the mathematical study of logical reasoning, proved fundamental theorems on continuous groups and on the fixed points of continuous mappings, and invented many of the basic notions of algebraic topology in the process of establishing the topological invariance of dimension.[2]

Brouwer earned a master's degree from the University of Amsterdam in 1904 with a thesis on continuous motions in real four-dimensional space. His doctoral dissertation [Brouwer, 1908], written under the delicate supervision of D. J. Korteweg and presented in 1907, was titled "On the Foundations of Mathematics" (*Over de grondslagen der wiskunde*) and divided into three parts: "The construction of mathematics," "Mathematics and experience" and "Mathematics and logic." Korteweg refused to accept some controversial material, but his student's critiques of the best-known mathematical philosophies of the time remained.

[1] A. Heyting, from the introduction to [Heyting, 1930a]; the translation is by A. S. Troelstra [1991].

[2] In one dimension, Brouwer's fixed point theorem implies that every continuous mapping of the real unit interval $[0, 1]$ into itself holds at least one point fixed. The invariance of dimension means that if there is a one-to-one continuous mapping from real n-space into real m-space, then $n \leq m$.

In Part I Brouwer built up the positive and negative integers and their arithmetic, then the rationals with their denumerable dense ordering, and finally a "measurable continuum," the result of constructing a *scale* of order type η (the order type of the rationals) on the raw one-dimensional continuum given by intuition.[3] In Part II he described the origin of scientific activity in the observation of "causal systems in time," disagreed with Kant by asserting that time is the only *a priori* element in science, and distinguished between mathematics ("an *action* and not a science") and metamathematics ("an *application* of mathematics, that is . . . an *experimental science*").[4],[5]

Part III began with the argument that mathematics is independent of and prior to logic, which depends on it; this was a direct and immediate challenge to the logicism of Frege and Russell, and also to the formalism of David Hilbert who was then the most influential logician (possibly the most influential mathematician) in Europe. Brouwer went on to criticize Cantor's set theory, accepting in principle every countable ordinal but rejecting Cantor's second number class (the class of *all* countable ordinals) as a completed totality because, being nondenumerable, it "cannot be mathematically constructed."[6] Finally, he argued in detail against logicism (which he attributed to Russell and Peano) and formalism.

The summary contained the following main points:

- "Mathematics is created by a free action independent of experience; it develops from a single aprioristic basic intuition, which may be called *invariance in change* as well as *unity in multitude*."[7]

[3] A linear ordering is of type η if and only if it is denumerable, has neither a smallest nor a largest element, and has the property that between any two elements there is another element.

[4] Arguing against Russell's defense of the Kantian *a priority* of space, Brouwer asserted that three-dimensional Euclidean geometry is just a six-parameter group which has proved useful to man in understanding and controlling the "phenomena of his world of experience."

[5] Hilbert adopted the name "metamathematics" for his program of attempting to establish, by finitary means, the consistency of formal systems representing logical and mathematical reasoning.

[6] Here Brouwer introduced the notion of a *denumerably unfinished set*, with "the following properties: we can never construct in a well-defined way more than a denumerable subset of it, but when we have constructed such a subset, we can immediately deduce from it, following some previously defined mathematical process, new elements which are counted to the original set. But from a strictly mathematical point of view this set does not exist as a whole; nor does its power exist; however we can introduce these words here as an expression for a known intention." Thus he rejected Cantor's platonism.

[7] Brouwer's footnote to the first point read: "The first act of construction has *two* discrete things thought together (also according to Cantor, Vortrag auf der Naturforscherversammlung in Kassel 1903); F. Meyer (Verhandl. internat. Math. Congr. Heidelberg 1904, p. 678) says that *one* thing is sufficient, because the circumstance that I think of it can be added as a second thing; this is false, for exactly this *adding* (i.e., setting it while the former is retained) *presupposes the intuition of two-ity*; only afterwards this simplest mathematical system is projected on the first thing and the *ego which thinks the thing*."

Meyer's claim, to which Brouwer objected, brings to mind the beginning of Dedekind's argument (p. 64 in the 1901 English translation of [Dedekind, 1888]) for the existence of infinite sets: "My own realm of thoughts, i.e., the totality S of all things, which can be objects of my thought, is infinite. For if s signifies an element of S, then is the thought s', that s can be object

- "Further, projecting mathematical systems on experience is also a free action, which shows itself efficient in the struggle for life; . . . "

- "A logical construction of mathematics, independent of the mathematical intuition, is impossible – for by this method no more is obtained than a linguistic structure, which irrevocably remains separated from mathematics . . . "

In April 1908 Brouwer had the honor of addressing the International Congress of Mathematicians in Rome, on the theory of Lie groups. The French preintuitionist Poincaré also spoke at the conference, on the future of mathematics.[8] In Part III of his doctoral dissertation Brouwer had already analyzed Poincaré's objections to logicism and to Cantor's set theory: "He blames mainly in logistics the petitio principii and in Cantorism the hypothesis of the actual infinite. But hereby he does not touch upon the heart of the question, which is situated deeper, namely in the confusion between the act of constructing mathematics and the language of mathematics."

That same year, in "The unreliability of the logical principles" (*De onbetrouwbaarheid der logische principes*), Brouwer asked,

> Is it allowed, in purely mathematical constructions and transformations, to neglect for some time the idea of the mathematical system under construction and to operate in the corresponding linguistic structure, following the principles of *syllogism*, of *contradiction* and of *tertium exclusum*, and can we then have confidence that each part of the argument can be justified by recalling to the mind the corresponding mathematical construction?

His answer was to argue for *modus ponens* (syllogism) and the law of contradiction but against the universal validity of the principle of excluded middle for properties of infinite systems, establishing a basis for what is now known as intuitionistic propositional logic. Observing that the law of contradiction justifies the double negation of the law of excluded middle, he added in a footnote: "Consequently, the theorems which are usually considered as *proved* in mathematics, ought to be divided into those that are *true* and those that are *non-contradictory*."

Following Korteweg's advice, Brouwer spent the next few years solving fundamental problems in topology. His inaugural lecture as an unpaid *privatdocent* or lecturer, given in 1909, was on the foundations of geometry and topology and included his view that "the *intuition of time* or *intuition of two-in-one*" lies behind all mathematical endeavor; while the nineteenth-century discovery of projective and other non-Euclidean geometries refuted Kant's belief in the *a priority* of (Eu-

of my thought, itself an element of S." In the preface Dedekind asserted that "numbers are free creations of the human mind."

[8] Dale M. Johnson, "Brouwer's coming of age as a topologist," page 65.

clidean) space.[9] His "On *analysis situs*," with ingenious topological counterexamples, and two papers on perfect sets quickly followed.

In 1912 these efforts were rewarded with a permanent appointment as Professor of Mathematics at the University of Amsterdam. An English translation of Brouwer's inaugural lecture, "Intuitionism and formalism," appeared in 1913 in the *Bulletin* of the American Mathematical Society. Here Brouwer used the term "neo-intuitionism" to refer to his own philosophy, distinguishing it from the old-style intuitionism which accepted Kant's belief in the *a priority* of both space and time. He objected strongly to the formalist position, which he identified with the belief that a mathematical theory can be justified by separating its content from its form (as in Peano's formal system for first-order arithmetic), and then proving the impossibility of deriving both a sentence (thought of as a string of symbols) and its negation from the axioms of the theory using formal logic.

The professorship gave him the freedom to continue developing and refining intuitionistic mathematics and mathematical philosophy, most remarkably in [Brouwer, 1918] and [Brouwer, 1919], as well as the stature to profoundly influence the ongoing foundational discussion which had begun before the turn of the century and was still going strong in 1930. It also brought him students, of whom Arend Heyting (1898–1980) was the one most devoted to intuitionism. Apparently realizing the desirability of communicating his own ideas in a way that would be comprehensible to the general mathematical community, Brouwer suggested that Heyting take as his doctoral subject the axiomatization and formalization of intuitionistic logic and parts of mathematics. Heyting responded with enthusiasm, and in 1925 he defended his dissertation, on the axiomatization of intuitionistic projective geometry. Related articles appeared in the *Mathematische Annalen* in 1927.[10]

In 1923 ([Brouwer, 1923], in Dutch) Brouwer published "Intuitionist splitting of the fundamental notions of mathematics," stressing the distinction between the classical meaning of negation as falsity and Brouwer's stronger interpretation of negation as absurdity. A German version [Brouwer, 1924a] and addendum [Brouwer, 1927a] followed. W. P. van Stigt points out [1998, p. 275] that this paper marked a radical departure for Brouwer, as it contained the statement and proof of a purely logical theorem: "Absurdity-of-absurdity-of-absurdity is equivalent with absurdity." It also demonstrated the use of negation and double negation as tools for defining new relations between mathematical objects; for example, beginning with the constructively defined relations of *coincidence* and *apartness* between points of the plane, Brouwer first noted that coincidence was equivalent to the absurdity of apartness, and then defined the new concept of *deviation* as the

[9] In the same lecture Brouwer traced the beginning of the recognition that only the intuition of time need be considered as *a priori*, back to Descartes' seventeenth-century invention of analytic geometry. And from the author's note to the 1919 reprint: "If this lecture were given to-day, it would certainly mention ... the 'general theory of relativity', contrived after 1909 ..." (see the English translation in [Brouwer, 1975], p. 120).

[10] Alongside his logical and philosophical work, Heyting continued making technical contributions to intuitionistic mathematics until his retirement. See A. S. Troelstra's appreciation and complete bibliography [Troelstra, 1968] of Heyting's work.

absurdity of coincidence (hence the absurdity of the absurdity of apartness). Intuitionistically, deviation did not imply apartness; but by the theorem just quoted, the absurdity of deviation was equivalent to coincidence.

In 1926 R. Wavre published a comparison [Wavre, 1926] of classical and "empirist" logical reasoning; the common principles were those he considered to belong to "Brouwerian logic."[11] A lively discussion followed in print, with contributions by Lévy, Barzin and Errera, Glivenko, Heyting, Borel and Church, among others.[12] Glivenko [1928] incompletely formalized intuitionistic propositional logic; his second version [Glivenko, 1929] included two axioms proposed to him by Heyting, who would later credit Glivenko with simplifying his own axiomatization.[13]

Heyting's brief 1930 article "On intuitionistic logic" [Heyting, 1930] began by endorsing Glivenko's formalization of intuitionistic propositional logic, and went on "to examine how the logic conceived of by Mr. Glivenko is complicated by the addition of the idea of provability Mr. Levy has brought into discussion and that he unduly identifies with the Brouwerian assertion." Instead, wrote Heyting, *the Brouwerian assertion of* p (written $\vdash p$) is "It is known how to prove p. ... The idea of an existence of mathematical entities outside our minds must not enter into the proofs." Then, introducing a new formal symbol $+$ for "is provable," he examined in detail the relation between Brouwerian negation and $+$ and observed: "In order to be able to assert $\vdash \neg + p$, it suffices to reduce to a contradiction the supposition that one can find a construction proving p." Thus it could happen that for some p "one would have at the same time $\vdash \neg + p$ and $\vdash \neg\neg p$. One should abandon every hope of ever settling the question; the problem would be unresolvable."[14]

In 1928 Hilbert and Ackermann's "Principles of Mathematical Logic" (*Grundzüge der theoretischen Logik* [Hilbert and Ackermann, 1928]) appeared, with careful axiomatizations of classical propositional and predicate logic; proofs of the completeness and consistency of the first; and a consistency proof and a clear statement of the completeness problem for the second.[15] One year later Gödel earned his doc-

[11] Wavre accepted $(A \supset B) \supset (\neg B \supset \neg A)$ as empirically correct, observing that the converse holds empirically only if B is a negation. He did not assert Heyting's axiom $\neg A \supset (A \supset B)$.

[12] Wavre, Borel and Lévy exchanged views in the *Revue de Métaphysique et de Morale*. Barzin and Errera, Glivenko, and Heyting published in the *Bulletin* of the Belgian Royal Academy. Church [1928], in English, criticised Barzin and Errera's [1927]. Details and some translations are in [Mancosu, 1998]. The history can be found in [Hessling, 2003] and [Mancosu *et al.*, 2009].

[13] Like Wavre, Glivenko at first omitted the intuitionistic principle that a contradiction implies every statement; also missing from his first attempt was $A \supset (B \supset A)$.

[14] Quotations are from the translation of Heyting [Heyting, 1930] by Amy L. Rocha, [Mancosu, 1998, p. 306ff], with slightly different symbolism. The example Heyting used of a possibly unresolvable mathematical question was the Goldbach conjecture, that every even integer (greater than 2) can be expressed as the sum of two prime numbers. While this conjecture remains unresolved, it was not well chosen because its logical form is such that any proof of its double negation would easily yield, by intuitionistic arithmetic, a proof of the conjecture itself. Since Heyting asserted that in general "$\vdash +p$ has exactly the same meaning as $\vdash p$," not both $\vdash p$ and $\vdash \neg + p$ can hold.

[15] The first three editions of Hilbert and Ackermann's very influential textbook were devoted

torate from the University of Vienna with a proof of the completeness of classical predicate logic.[16]

Finally, in 1930 Heyting's three fundamental articles "The formal rules of intuitionistic logic" [Heyting, 1930a] and "The formal rules of intuitionistic mathematics" II and III [Heyting, 1930b; Heyting, 1930c] appeared in German in the *Sitzungsberichte* of the Prussian Academy of Sciences. The first of these formalized intuitionistic propositional logic, the second presented formal systems for intuitionistic first-order predicate logic and arithmetic, and the third contained a partial axiomatization of intuitionistic analysis and Brouwer's theory of species (his substitute for classical set theory). Heyting's work, which will be discussed in detail here, was partly anticipated by Glivenko and also by A. Kolmogorov, whose axiomatization [Kolmogorov, 1925] of the intuitionistic propositional logic of implication and negation ("minimal logic") and part of intuitionistic first-order predicate logic was published in 1925 in Russian; it does not appear to have been noticed in the West.[17]

In the 1930s Kurt Gödel and Gerhard Gentzen independently observed that classical logic and arithmetic could be effectively translated into their intuitionistic subtheories. Gödel stated, and Gentzen proved, the disjunction and existence properties for intuitionistic formal logic and arithmetic. A decade later Brouwer's work caught the attention of the American logician Stephen Kleene who eventually, with his student Richard Vesley, developed in [Kleene and Vesley, 1965] an efficient axiomatization and a consistency proof for intuitionistic analysis, clarifying its relation to the classical theory. In 1958 Gödel proved the consistency of classical analysis using primitive recursive functions of higher type (the "Dialectica interpretation" [Gödel, 1958]; cf. [Avigad and Feferman, 1998; **?**]). Four years later Clifford Spector, another Kleene student, used bar recursive functionals for the same purpose in [Spector, 1962], extending to higher types one of the key principles of intuitionistic analysis.

In 1973 Heyting's student Anne S. Troelstra edited (and mostly wrote) the influential survey [Troelstra, 1973], which was followed in the 1980s by Michael Beeson's [Beeson, 1985] and Troelstra and Dirk van Dalen's [Troelstra and van Dalen, 1988]. Many other logicians have contributed to the metamathematical study of intuitionistic mathematics and logic, among them P. Aczel, E. W. Beth, M. Bezem, J. Diller, M. Dummett, A. G. Dragalin, M. Fitting, M. Fourman, H. Friedman, S. Ghilardi, R. Grayson, S. Hayashi, G. F. van der Hoeven, W. A. Howard, R. Iemhoff, J. J. de Iongh, D. H. J. de Jongh, G. Kreisel, S. Kripke,

entirely to classical logic. In the fourth German edition, published in 1959, Ackermann added optional sections on intuitionistic logic and strong negation.

[16]For a full discussion of the classical completeness problem, its history and solution, see the article in this volume titled "Gödel's logic" [**?**]. There, van Atten and Kennedy present evidence that Gödel was influenced by two lectures Brouwer gave in Vienna in March, 1928. They also quote a passage from Gödel's doctoral dissertation, crediting Brouwer with stressing that the consistency of a formal system did not in itself imply the existence of a model.

[17]Mancosu, Zach and Badesa [Mancosu *et al.*, 2009] includes an analysis of the content of [Kolmogorov, 1925].

M. Krol, H. Läuchli, D. Leivant, V. Lifschitz, M. Löb, I. Loeb, H. Luckhardt, P. Martin-Löf, D. McCarty, G. Mints, I. Moerdijk, J. Myhill, D. Nelson, J. van Oosten, H. Rasiowa, W. Ruitenberg, B. Scarpellini, A. Scedrov, K. Schütte, D. Scott, P. Scowcroft, C. A. Smorynski, H. C. M. de Swart, A. S. Troelstra, W. Veldman, A. Visser, F. Waaldijk, S. Weinstein and the author.[18]

Brouwer's chair at the University of Amsterdam was held after his death by Heyting, Troelstra and de Jongh in succession. The tradition of doing intuitionistic mathematics in the spirit of Brouwer has been continued in Nijmegen in the Netherlands by Wim Veldman, a student of J. J. de Iongh, who was a student of Brouwer. Two alternative schools of mathematics based on intuitionistic logic are the Russian recursive mathematics of A. A. Markov and his followers, and Errett Bishop's cautious constructive mathematics which is still under development by Douglas Bridges, Fred Richman, Hajime Ishihara and many others. Intuitionistic logic has found modern applications in computer science and category theory.

2 INTUITIONISTIC PROPOSITIONAL LOGIC

2.1 Brouwer's View of Negation as Absurdity

Brouwer objected equally to the laws of excluded middle and double negation because he interpreted negation not as falsity, but as absurdity. Thus "p or not p" meant "either p is correct or p is contradictory ;" while "if not not p, then p" meant that the absurdity of the absurdity of p (in other words, the consistency of p) implies that p is correct.[19] In [Brouwer, 1925] Brouwer thanked Bernays for pointing out that the general law of excluded middle is derivable from the general law of double negation; the converse observation had already appeared in [Brouwer, 1908], where Brouwer equated the general law of excluded middle with the (untenable) *a priori* position that every mathematical question is solvable.

Kolmogorov [1932] observed that Brouwer's view transforms a negative assertion into a positive one: "Thus the negation of a is transformed into an *existential proposition*: 'There exists a chain of logical inferences that, under the assumption of the correctness of a, leads to a contradiction.'" Under material implication, which Heyting explicitly accepted, every statement is derivable from any contradiction. Thus, given a particular contradictory assertion $\perp$, the negation of a can be defined as "if a then $\perp$," reducing negation to implication. Nevertheless, all the early intuitionistic formal systems include negation as a primitive symbol.

[18]Here the term "metamathematical" is used loosely to include the study, by constructive or classical means, of formal systems whose intended interpretation is intuitionistic. Brouwer did not long object in principle to formalizing parts of intuitionistic reasoning, but it is unlikely that he would have approved of using classical methods to study the resulting formal systems.

[19]W. P. van Stigt singles out [Brouwer, 1923], in which Brouwer proves that "absurdity of absurdity of absurdity is equivalent with absurdity," as evidence of his (temporary) willingness to study metamathematical questions. At that time Brouwer was also interested, with Mannoury, in the Significist philosophy which held that it was possible to improve communication by refining language.

2.2 Heyting's Propositional Calculus

From a historical perspective the most significant formalization of intuitionistic propositional logic was the one published by Heyting in 1930 [Heyting, 1930a]. We present it here using Heyting's notation, but omitting the " $\vdash$ " preceding each axiom. After observing that the intuitionistic logical connectives $\supset$, $\wedge$, $\vee$, $\neg$ are independent of one another (unlike the classical connectives), he proposed as rules of inference:

1.2. If a and b are correct formulas, then $a \wedge b$ is also correct.

1.3. If a and $a \supset b$ are correct formulas, then b is also correct.

and as axiom schemata:

2.1. $\cdot\, a \supset a \wedge a$.

2.11. $\cdot\, a \wedge b \supset b \wedge a$.

2.12. $\cdot\, a \supset b \cdot \supset \cdot\, a \wedge c \supset b \wedge c$.

2.13. $\cdot\, a \supset b \;\cdot \wedge \cdot\; b \supset c \cdot \supset \cdot\, a \supset c$.

2.14. $\cdot\, b \supset \cdot\, a \supset b$.

2.15. $\cdot\, a \wedge \cdot\, a \supset b \cdot \supset b$.

3.1. $\cdot\, a \supset a \vee b$.

3.11. $\cdot\, a \vee b \supset b \vee a$.

3.12. $\cdot\, a \supset c \;\cdot \wedge \cdot\; b \supset c \cdot \supset \cdot\, a \vee b \supset c$.

4.1. $\cdot\, \neg a \supset \cdot\, a \supset b$.

4.11. $\cdot\, a \supset b \;\cdot \wedge \cdot\; a \supset \neg b \cdot \supset \neg a$.

In each rule or axiom the a, b, c are *metavariables* in the sense that each one may be replaced, throughout the statement of the rule or axiom, by any well-defined formula.

Heyting defined logical equivalence by

2.01. $\cdot\, a \supset\subset b \cdot =_D \cdot\, a \supset b \;\cdot \wedge \cdot\; b \supset a$.

and proceeded to sketch proofs for seventy formal theorems. As an example, consider[20]

2.2. $\vdash \cdot\, a \wedge b \supset a$.

Proof. [2.14] $\vdash \cdot\, a \supset \cdot\, b \supset a \;:\supset:$

[20] In Heyting's version the last "a" of the proof was incorrectly printed as "b."

$$[2.12] \vdash \cdot\, a \wedge b \supset \cdot\, b \supset a \;\cdot \wedge\, b \cdot \supset [2.15]\; a\ .$$

Observe that the last step involves an implicit use of axiom 2.11; the applications of rule 1.3 are also implicit; but the argument is evidently sound, provided the postulates are correct. We single out a few additional numbered results, omitting the proof-sketches (which depend on intermediate theorems).[21]

2.22. $\vdash \cdot\, a \wedge b \supset b$.

2.26. $\vdash \cdot\, b \supset \cdot\, a \supset a \wedge b$.

4.2. $\vdash \cdot\, a \supset b \cdot \supset \cdot\, \neg b \supset \neg a$.

4.3. $\vdash \cdot\, a \supset \neg\neg a$.

4.31. $\vdash \cdot\, \neg a \supset \neg\neg\neg a$.

4.32. $\vdash \cdot\, \neg\neg\neg a \supset \neg a$.

4.45. $\vdash \cdot\, a \vee \neg a \cdot \supset \cdot\, \neg\neg a \supset a$.

4.46. $\vdash \cdot\, \neg a \vee b \supset \cdot\, a \supset b$.

4.47. $\vdash \cdot\, a \vee b \supset \cdot\, \neg a \supset b$.

4.8. $\vdash \cdot\, \neg\neg(a \vee \neg a)$.

4.83. $\vdash \cdot\, a \vee \neg a \supset \neg b \cdot \supset \neg b$.

4.9. $\vdash \cdot\, a \supset b \cdot \supset \neg(a \wedge \neg b)$.

4.91. $\vdash \cdot\, a \vee b \supset \neg(\neg a \wedge \neg b)$.

4.92. $\vdash \cdot\, a \wedge b \supset \neg(\neg a \vee \neg b)$.

To show each of his axioms independent of the other ten, Heyting provided eleven interpretations in the style of Bernays, using finite or countably infinite truth tables. To show axiom 2.15 is independent, for example, he used four truth values 0, 1, 2, 3 (of which only 0 represents "correct") with the tables

$\supset$	0	1	2	3
0	0	0	0	0
1	1	0	1	1
2	2	0	0	0
3	2	0	2	0

$\wedge$	0	1	2	3
0	0	1	2	2
1	1	1	1	1
2	2	1	2	2
3	2	1	2	2

$\vee$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	2	2
3	0	3	2	3

$\neg$	0	1	2	3
	1	0	1	1

When $a,\ b,\ c$ independently take arbitrary values in $\{0, 1, 2, 3\}$ every axiom but 2.15 takes the value 0 . But when a is 0 and b is 3 then $a \wedge \cdot\, a \supset b$ takes the value $0 \wedge 2 = 2$, so $a \wedge \cdot\, a \supset b \cdot \supset b$ is $2 \supset 3 = 2 \neq 0$.

[21] Heyting credited Glivenko with the formal proofs of 4.3 and 4.8, and with the statement of 4.83.

The requirement that the two rules of inference lead from correct formulas to correct formulas imposed restrictions on the interpretations. Heyting observed that for rule 1.2 the condition $0 \wedge 0 = 0$ was necessary and sufficient; while for rule 1.3 it was enough that $0 \supset a = 0$ only when $a = 0$ (so the first column of the table for $\supset$ must have only one 0 representing the value of $0 \supset 0$). However, he failed to notice that rule 1.2 was derivable from the other postulates. To see this, suppose a and b are correct. Then so are $b \supset a$ (by 2.14 with rule 1.3) and $b \wedge b$ (by 2.1 with rule 1.3). But $b \supset a \cdot \supset \cdot b \wedge b \supset a \wedge b$ is an instance of 2.12, so $a \wedge b$ follows by two applications of rule 1.3.

Finally Heyting provided a three-valued interpretation for which all the axioms and rules are correct, but not $\neg\neg a \supset a$ (and hence not $a \vee \neg a$, by theorem 4.45 stated above). Note that for Heyting "correctness" was absolute: "The formula $a \supset b$ means in general: 'If a is correct, then b is also correct'," [22] from which it follows that if b is correct, so is $a \supset b$ (the sense of axiom 2.14).

As his main sources Heyting listed Peano's *Formulaire des Mathématiques*, Russell and Whitehead's *Principia Mathematica*, Hilbert and Ackerman's *Grundzüge der theoretischen Logik*, and two articles: Bernays' "Axiomatische Untersuchung des Aussagenkalküls der 'Principia Mathematica' " (Math. Zeitschrift 25 (1926) p. 305) and Glivenko's "Sur la logique de M. Brouwer" (Bulletins de l'Academie royale de Belgique 5^e serie vol. XIV (1928) p. 225). Glivenko's article deserves a closer look, as it both anticipated and was influenced by Heyting's fomalization.

2.3 Glivenko's Contribution

Inspired by Brouwer's [1923], Valerii Glivenko published two notes in French in the *Académie Royale de Belgique* in 1928 [Glivenko, 1928] and 1929 [Glivenko, 1929]. The first contained an incomplete axiomatization of intuitionistic propositional logic with *modus ponens* as its only rule of inference, and a formal proof of $\neg\neg(a \vee \neg a)$ ("a remarkable theorem of Mr. Brouwer" [Glivenko, 1929]).

The second provided four additional axioms, including Heyting's axioms 2.14 and 4.1 (which Glivenko justified intuitively as consequences of the intuitionistically acceptable principle $a \vee \neg b \cdot \supset \cdot b \supset a$).[23] On the basis of his new axiomatization, which he claimed to be complete for intuitionistic propositional reasoning, Glivenko proved two metatheorems (known today as "Glivenko's Theorems") relating intuitionistic and classical logic. In his own words:[24]

1. "If a certain expression in the logic of propositions is provable in classical logic, it is the falsity of the falsity of this expression that is provable in Brouwerian logic."

[22] p. 44 of the German original.

[23] In a footnote Glivenko gave Heyting credit for convincing him of the necessity of these two axioms, and referred to Heyting's "memoir, which will soon appear in the *Mathem. Annalen*." In fact Brouwer, who had fallen out of favor with Hilbert, was no longer an editor of that journal and Heyting's formalization appeared instead in the Proceedings of the Prussian Academy of Sciences as [Heyting, 1930a; Heyting, 1930b] and [Heyting, 1930c]; cf. [van Dalen, 1990].

[24] The translation is by Amy L. Rocha, [Mancosu, 1998, p. 301].

2. "If the falsity of a certain expression in the logic of propositions is provable in classical logic, that same falsity is provable in Brouwerian logic."

To prove the first theorem he used *modus ponens* with $a \supset \neg\neg a$ (already proved in [Glivenko, 1928]) to derive the double negation of each of his axioms from the axiom itself, then verified that $\neg\neg b$ is derivable in his augmented intuitionistic system from $\neg\neg a$ and $\neg\neg(a \supset b)$. The second was an easy corollary of the first, using $\neg\neg\neg a \supset \neg a$ (also proved in [Glivenko, 1928]).

Glivenko did not address the question whether all his axioms were necessary, but each is provable in Heyting's formalism. Conversely, each of Heyting's axioms is provable from Glivenko's, and Heyting's rule 1.2 follows from Glivenko's axioms by *modus ponens* (which is Heyting's rule 1.3). From this point of view the two formal systems are equivalent, although each has advantages and disadvantages. In [Kleene, 1952], as we shall see, Kleene made a coherent selection of axioms from their union.

2.4 Kolmogorov's Problem Interpretation

In 1932 Kolmogorov [1932] provided a kind of semantics for intuitionistic logic, as an alternative to Heyting's emphasis on proofs. Taking the notions of (mathematical) "problem" and "solution of a problem" as already understood, he proposed to attach to every expression of the form $\vdash p(a, b, c, \ldots)$ the problem "to give a general method for the solution of $p(a, b, c, \ldots)$ for any single choice of problems $a, b, c, \ldots$." Here $p(a, b, c, \ldots)$ is a propositional formula with only the atoms $a, b, c, \ldots$ and the propositional connectives have the following interpretation:

> If a and b are two problems, then $a \wedge b$ designates the problem "to solve both a and b," while $a \vee b$ designates the problem "to solve at least one of the problems a and b." Furthermore, $a \supset b$ is the problem "to solve b provided that the solution for a is given" or, equivalently, "to reduce the solution of b to the solution of a." ... Correspondingly, $\neg a$ designates the problem "to obtain a contradiction provided that the solution of a is given."

Kolmogorov thus transformed the logical rules and axioms of Heyting's propositional logic into the postulates of an "elementary calculus of problems."[25]

These ideas, which Kolmogorov extended to predicate logic (see the next section), anticipated Kleene and Nelson's 1945 realizability interpretation [Kleene, 1945; Nelson, 1947] of intuitionistic arithmetic. Together with Heyting's proof interpretation they form the basis of what is often called the "Brouwer-Heyting-Kolmogorov (B-H-K) explication" of intuitionistic logic.

[25]The quotations are from Paolo Mancosu's translation, [Mancosu, 1998], [Kolmogorov, 1932, pp. 328–334].

3 INTUITIONISTIC FIRST-ORDER PREDICATE LOGIC

3.1 The Intuitionistic Universal and Existential Quantifiers

Brouwer accepted universal quantification over the natural numbers, interpreting the statement that every n has a certain property as an incomplete communication of a construction which, applied in a uniform manner to each natural number n as it is generated, verifies the property for that n. The statement that there exists an m with a certain property means that a particular number m with the property has been constructed, or that a method for constructing such a number is known. Kolmogorov similarly interpreted a universal statement as a problem requiring a general method for its solution, and an existential statement as the problem of finding a witness. Thus under the B-H-K interpretation the phrase "for every n there is an m" affirms the construction of an effective operation which assigns to each n an appropriate m; this is why the axiom of countable choice is generally considered to be intuitionistically correct.

Quantification over more complicated domains is meaningful intuitionistically only if the elements of the domain are legitimate mathematical entities and the structure of the domain admits uniform effective operations. This interpretation clearly distinguishes intuitionistic quantification from the classical case where each domain is assumed to consist of preexisting objects which can be examined simultaneously for the presence or absence of a given property, and the existential quantifier does not entail the production of a witness.

3.2 Heyting's Predicate Calculus With Identity

Continuing the work he began in [Heyting, 1930a] (§2.2 above), in the introduction to [Heyting, 1930b] Heyting wrote:[26]

> To enable the further decomposition into subject and predicate, we follow the path of functional calculus which was first laid down by Russell and Whitehead, and was recently lucidly presented and significantly improved by Hilbert and Ackermann in their book. Above all, the concept of mathematical object must be explained. The real construction of mathematics begins immediately with the treatment of objects (natural numbers, sets, etc.) For the formal treatment, however, it turns out to be convenient to assume in advance general axioms about the relations between objects and about formulas which contain variables for objects.. . .
>
> The fundamental relation between individuals is $p \stackrel{\equiv}{=} q$, " p is the same object as q." . . . The symbol $=$ will be used for numerical equality and similar concepts (e.g. cardinal equality for sets); this relation can hold

[26] Unless otherwise noted, all translations from the German of [Heyting, 1930b; Heyting, 1930c] are my own.

> between objects which are not identical; ... We also need a special symbol for mathematical identity, which must be defined for each sort of mathematical objects ... For mathematical identity we choose the symbol $\equiv$.[27]

He attributed to the formula $p \stackrel{\equiv}{=} p$ a strong sense according to which "we can translate 6.1. $\vdash\vdash 1 \stackrel{\equiv}{=} 1$ by '1 is an object.'"[28] He used the notation $\left(\begin{smallmatrix} p \\ x \end{smallmatrix} \right) a$ for "'the expression which results from a by replacing the variable x throughout, everywhere it appears, by the term (*Zeichenzusammenstellung*) p.'" Building on Hilbert and Ackermann's [Hilbert and Ackermann, 1928] and his own [Heyting, 1930a], Heyting developed the formal language of his predicate calculus using $a, b, c, \ldots$ for sentential variables (*Aussagenveränderliche*) and $p, q, r, \ldots$ for object variables (*Gegenstandveränderliche*), with the relation symbols $\stackrel{\equiv}{=}$, ε, the function application symbol ', and "objects (*Gegenstände*) I, N, seq" and specifying that the "quantifiers $\left(\begin{smallmatrix} p \\ x \end{smallmatrix} \right), (x), (Ex)$, where x is an object variable, make x into a *bound variable* in the part of the formula they influence. Such [an x] can only be replaced by another object variable."

His "rules of operation" included an inductive definition of "expression" (i.e. well-formed formula) extending the formation rules for $\supset, \wedge, \vee$ and $\neg$ stated more casually in [Heyting, 1930a], but for one clause of this definition the concept "correct formula" — which itself depended inductively on the axiom schemata and rules of inference, including those of the propositional calculus — was required by his use of $\stackrel{\equiv}{=}$ to confer legitimacy on terms; see rule 5.32 below. Heyting did not define "term" but the usage made clear that object variables and constant objects such as I were terms. We list his key new rules, omitting a few obvious conditions on substitution in expressions but including the formation rules and the rules of inference 5.4–5.7:[29]

5.3. Every correct (*richtige*) formula is an expression (*Ausdruck*).

5.32. From an expression another expression results, if one replaces a propositional variable throughout by the same expression, or an object variable throughout by the same term p, for which $p \stackrel{\equiv}{=} p$ is a correct formula.

---

[27] In each mathematical application, $=$ and $\equiv$ must be defined for each sort of object in terms of $\stackrel{\equiv}{=}$. In arithmetic there is no difference.

[28] This use of $\stackrel{\equiv}{=}$ anticipated by decades a similar use of $=$ in the logic of partial terms, cf. [Scott, 1979; Beeson, 1985].

[29] Richard Zach has observed an anomaly in Heyting's definition of "expression." Rule 5.7 (with 5.3) allows $(y) \cdot \left(\begin{smallmatrix} y \\ x \end{smallmatrix} \right) A \supset\subset A$ to be an expression if x occurs (only) bound in A, but then no rule guarantees that e.g. $\left(\begin{smallmatrix} y \\ x \end{smallmatrix} \right) A$ is an expression. The gap could be filled by extending the "and conversely" in rule 5.34 to rule 5.36 and adding a rule to the effect that if $\left(\begin{smallmatrix} p \\ x \end{smallmatrix} \right) A$, $(x)\, A$, or $(Ex)\, A$ is an expression, so is A.

5.34. If A is an expression, so is $\neg A$, and conversely.

5.35. If A is an expression in which x does not appear as a bound variable, then $\left(\begin{array}{c} p \\ x \end{array} \right) A$, $(x)\, A$, $(Ex)\, A$ are expressions.

5.36. If A and B are expressions, so are $A \supset B$, $A \wedge B$ and $A \vee B$.

5.4. From a correct formula another correct formula results, if one replaces a propositional variable throughout by the same expression, or an object variable throughout by a term p, for which $p \overset{\equiv}{=} p$ is a correct formula.

5.5. If $a \supset b$ and $(p) \cdot \left(\begin{array}{c} p \\ x \end{array} \right) a \supset\subset a$ are correct formulas, then $a \supset (x)\, b$ is a correct formula.

5.6. If $a \supset b$ and $(p) \cdot \left(\begin{array}{c} p \\ x \end{array} \right) b \supset\subset b$ are correct formulas, then $(Ex)\, a \supset b$ is a correct formula.

5.7. If in the expression A the variable x appears only bound, then $(y) \cdot \left(\begin{array}{c} y \\ x \end{array} \right) A \supset\subset A$ is a correct formula.

As axiom schemata, in addition to those of intuitionistic propositional logic extended to the predicate language, Heyting proposed the following. First, the axioms for equality:

6.1. $1 \overset{\equiv}{=} 1$.

6.11. $p \overset{\equiv}{=} p \wedge q \overset{\equiv}{=} q \supset p \overset{\equiv}{=} q \vee \neg(p \overset{\equiv}{=} q)$.

6.12. $q \overset{\equiv}{=} p \supset p \overset{\equiv}{=} q$.

The next group contained axioms asserting that the replacement operator commutes with the propositional connectives, and also

6.2. $\left(\begin{array}{c} x \\ x \end{array} \right) a \supset\subset a$.

6.21. $\left(\begin{array}{c} \left(\begin{array}{c} q \\ x \end{array} \right) p \\ x \end{array} \right) a \supset\subset \left(\begin{array}{c} q \\ x \end{array} \right) \left(\begin{array}{c} p \\ x \end{array} \right) a$.

6.26. $\cdot\, p \overset{\equiv}{=} q \supset \cdot \left(\begin{array}{c} p \\ x \end{array} \right) a \supset \left(\begin{array}{c} q \\ x \end{array} \right) a$.

Heyting then introduced a notation $g(\overline{x})$ for "g does not contain x," with the definition

6.01. $\cdot\ g(\overline{x}) =_D \cdot\ (p)\begin{pmatrix} p \\ x \end{pmatrix} g \equiv g \cdot \vee \cdot\ (p) \cdot \begin{pmatrix} p \\ x \end{pmatrix} g \supset\subset g\ .$

Without commenting on the peculiarity of the disjunctions obtained by literally replacing g throughout this definition by a term or expression, he remarked: "... The first alternative in 6.01 holds, when g represents an object; the second, if g is an expression. The introduction of this symbol will make possible the distinction between function and sentence variables." This aspect of Heyting's formalism illustrates its incompatibility with modern treatments like that of Kleene [Kleene, 1952, p. 82], which will be considered later.

Six quantifier axioms completed the predicate calculus:

6.3. $(x)\ a \supset \begin{pmatrix} p \\ x \end{pmatrix} a\ .$

6.31. $(x)\ a \supset (y)\begin{pmatrix} y \\ x \end{pmatrix} a\ .$

6.32. $\cdot\ (y) \cdot\ y \equiv y \supset \begin{pmatrix} y \\ x \end{pmatrix} a \cdot \supset (x)\ a\ .$

6.5. $\begin{pmatrix} p \\ x \end{pmatrix} a \supset (Ex)\ a\ .$

6.51. $\cdot\ a\ (\overline{x}) \supset \cdot\ (Ex)\ a \supset (x)\ a\ .$

6.52. $\cdot\ (x) \cdot a \supset b \cdot \supset \cdot\ (Ex)\ a \supset (Ex)\ b\ .$

The thirty five formal theorems Heyting proved on this basis included

6.77. $\vdash (Ex) \neg \neg a \supset \neg \neg (Ex)\ a\ .$

6.78. $\vdash \neg \neg (x)\ a \supset (x) \neg \neg a\ .$

He remarked that neither converse is provable, observing "For 6.77 this follows easily from the meaning of (Ex)." For 6.78 he used a *Brouwerian counterexample*, the set A of all infinitely proceeding sequences of digits 1, 2, such that once a 2 is chosen, only 2 may be chosen thereafter. To each (positive) natural number he assigned an element of A as follows. The number 1 is associated with the constant sequence $1111\ldots$, 2 with the constant sequence $2222\ldots$, 3 with $1222\ldots$, 4 with $1122\ldots$, and so on. Letting the object variable x range over the elements of A and taking a to mean "x is assigned to a natural number," Heyting argued that $(x) \neg \neg a$ held for this interpretation by the definition of A, while $\neg (x)\, a$ held by the nonclassical form of Brouwer's fan theorem, see §7 below. "However by 4.521 $(x) \neg \neg a \wedge \neg (x)\, a$ is incompatible with the converse of 6.78."[30]

[30] In [Brouwer, 1925, p. 256], from which Heyting said he took this example, Brouwer drew a stronger conclusion: that the set described is not effectively enumerable (in the sense that *no* correlation of distinct natural numbers to all of its elements is possible).

4 INTUITIONISTIC ARITHMETIC

In the remainder of [Heyting, 1930b] Heyting applied intuitionistic predicate logic to arithmetic. The language he chose was so arcane that in the interest of comprehension it will not be used here except in this brief discussion from the historical viewpoint.

4.1 Heyting's Arithmetic

From the primitive symbols ε and ‘, with the intended interpretations of membership and function application respectively, he first constructed formal definitions (or abbreviations) corresponding to the concepts "x is a single-valued function from p to q" and "s belongs to the image of r under x," then defined mathematical identity between two functions from p to q extensionally in terms of $\overline{\overline{=}}$ on the range. Using the constant symbol N to represent the set of positive natural numbers, and seq for the successor operation, he added axioms asserting that 1 is a natural number, successor is a one-to-one function from numbers to numbers, and 1 is not a successor; also the axiom schema of mathematical induction. He defined numerical equality $=$ and mathematical identity $\equiv$ between numbers to be the same as $\overline{\overline{=}}$. Then he gave the primitive recursive defining equations for $+$ and (quantifier-free) explicit definitions for $-$, $<$ and $>$, and developed the quantifier-free part of this theory. Among the formal theorems he proved was

10.524. $\vdash p, q \varepsilon N \supset p = q \vee p > q \vee p < q$.

Heyting did not draw the conclusion that *every* expression A involving only 1, *seq*, $=$, $+$, $-$, $>$, $<$ and the propositional connectives is formally decidable when all individual variables range over N, although he could have shown this by a straightforward induction on the logical form of A. He did not add a symbol and definition for multiplication. That same year (1930), in the Proceedings of the First Congress of Mathematicians of the Slavic Countries held at Warsaw in 1929, M. Presburger [1930] published consistency and completeness proofs and a decision procedure for a formal system of classical first-order arithmetic with addition but without multiplication.[31]

What is now usually meant by "Heyting's arithmetic" is a first-order formalism with (at least) an individual constant 0, a unary function symbol S or $'$, binary function symbols $+$ and $\cdot$, and axioms in which the individual variables range implicitly rather than explicitly over the natural numbers. The importance of multiplication is that it permits the coding of finite sequences of numbers by numbers, and this in turn makes possible the coding of arithmetic within arithmetic; cf. §5.1 below. Troelstra [1973] included the axioms for all primitive re-

[31] Cf. Kleene's [1952, p. 204]. Presburger's article is in German but the title of the volume is in Polish. Kleene noted that Hilbert and Bernays [1934, pp. 359ff.] adapted Presburger's proof to a less cumbersome classical formalism, and Joan Ross verified the corresponding result for the intuitionistic subsystem of Presburger arithmetic .

cursive number-theoretic functions, making his version of Heyting's arithmetic an extension of Primitive Recursive Arithmetic **PRA** and a subtheory of "Heyting's arithmetic of finite types" $\mathbf{HA}^{\omega}$.

4.2 Kleene's Intuitionistic Arithmetic

Anticipating Kleene's formalization of intuitionistic analysis, to be discussed in §6 below, we now present his standard treatment of intuitionistic arithmetic. With minor changes in symbolism and numbering this is the formal system developed and studied in Kleene's fundamental 1952 textbook *Introduction to Metamathematics* [Kleene, 1952], which played the same role for logicians in the third quarter of the twentieth century as Hilbert and Ackermann's [1928] in the second. Kleene's language has countably many individual variables $a_1, a_2, a_3, \ldots$, an individual constant 0, a unary function symbol $'$, two binary function symbols $+$ and $\cdot$, and the binary predicate symbol $=$. There are no other predicate symbols. *Terms* and *prime formulas* are defined inductively as follows:

- 0 is a *term*.
- Each individual variable is a *term*.
- If s and t are *terms* then s', $(s + t)$ and $(s \cdot t)$ are *terms*.
- If s and t are *terms* then $(s = t)$ is a *prime formula*.

Every occurrence of a variable x in a term s or t is *free in* $(s = t)$. Parentheses and the symbol $\cdot$ may be omitted according to the usual mathematical conventions. *Formulas* are built from prime formulas using $\wedge, \vee, \supset, \neg, \forall, \exists$, and parentheses (omitted by the usual conventions when the scope of a logical symbol is clear without them, as Heyting omitted unnecessary delimiting dots). An occurrence of a variable x in a formula A is *bound* if it is within the scope of an existential or universal quantifier, otherwise *free*.

The only *rules of inference* are the logical rules R1 - R3 inherited from Kleene's treatment of propositional and predicate logic (compare Heyting's rules 1.3 and 5.5 - 5.7) :

R1. From A and $A \supset B$, conclude B.

R2. From $C \supset A(x)$ where x does not occur free in C, conclude $C \supset \forall x A(x)$.

R3. From $A(x) \supset C$ where x does not occur free in C, conclude $\exists x A(x) \supset C$.

The *axioms* are of three kinds: the logical axiom schemas X1 - X12, the mathematical induction schema X13, and the particular arithmetical axioms X14-X21 (from which the reflexivity, symmetry and transitivity of $=$ follow). For X11 and X12, $A(x)$ may be any formula and t any term free for x in $A(x)$. For X13, $A(x)$ may be any formula and x any variable. For X14 - X21 the a, b, c are three particular distinct individual variables (for example a_1, a_2, a_3), so these axioms

(unlike X13) are formulas rather than schemas. Their universal closures are provable, and the conclusion of X13 can be strengthened to $\forall x A(x)$ by R2 since x is not free in the hypothesis.

X1. $A \supset (B \supset A)$.

X2. $(A \supset B) \supset ((A \supset (B \supset C)) \supset (A \supset C))$.

X3. $A \supset (B \supset A \ \wedge \ B)$.

X4. $A \ \wedge \ B \supset A$.

X5. $A \ \wedge \ B \supset B$.

X6. $A \supset A \vee B$.

X7. $B \supset A \vee B$.

X8. $(A \supset C) \supset ((B \supset C) \supset (A \vee B \supset C))$.

X9. $(A \supset B) \supset ((A \supset \neg B) \supset \neg A)$.

X10. $\neg A \supset (A \supset B)$.

X11. $\forall x A(x) \supset A(t)$.

X12. $A(t) \supset \exists x A(x)$.

X13. $A(0) \ \wedge \ \forall x(A(x) \supset A(x')) \supset A(x)$.

X14. $(a' = b') \supset (a = b)$.

X15. $\neg(a' = 0)$.

X16. $(a = b) \supset ((a = c) \supset (b = c))$.

X17. $(a = b) \supset (a' = b')$.

X18. $(a + 0) = a$.

X19. $(a + b') = (a + b)'$.

X20. $(a \cdot 0) = 0$.

X21. $(a \cdot b') = (a \cdot b) + a$.

After developing an unambiguous coding in this theory for finite sequences of natural numbers, Kleene proved that symbols and defining equations for any number of primitive recursive functions could be conservatively added to the system. He went on to give a careful exposition of the negative results of Gödel, Church and Rosser for first-order intuitionistic number theory as well as for the classical theory which results when Heyting's axiom schema X10 is replaced by $\neg\neg A \supset A$.

We state these results in the next section; for details, and also for historical evidence of Brouwer's influence on Gödel, see van Atten and Kennedy's [**?**] in this volume.

Kleene's formalization of pure intuitionistic first-order predicate logic without equality is obtained by allowing only individual variables as terms, allowing as prime formulas only proposition letters and expressions formed from n-place predicate letters by filling the places with occurrences of variables, and omitting the arithmetical axioms X13 - X21. From now on we refer to this theory, based on rules R1 - R3 and axiom schemas X1 - X12, as "intuitionistic predicate logic" **IQC**. By "intuitionistic propositional logic" **IPC** we mean the subtheory of **IQC** with only proposition letters as prime formulas, only the rule R1 and only the axiom schemas X1–X10.[32]

5 PROOF THEORY OF INTUITIONISTIC LOGIC AND ARITHMETIC

Classical propositional and predicate logic, and a version of classical (Presburger) arithmetic, can be obtained from Heyting's formal systems simply by replacing axiom schema 4.1 by either the law of excluded middle or the law of double negation; then 4.1 becomes a theorem. Kleene's formalisms have the analogous property. Thus intuitionistic logic and arithmetic are proper subtheories of the corresponding classical theories.

Classical propositional logic can be formalized using only two logical connectives ($\wedge$ and $\neg$, for example), from which the others are definable. Classical predicate logic needs only one quantifier (for example, $\forall$) from which the other is definable using negation.[33] The intuitionistic logical language is thus more expressive than the classical, allowing finer distinctions. But is intuitionistic logic essentially weaker than classical logic? This question was answered in stages, and the answer is a firm "no."

5.1 Equiconsistency and the Negative Translations

Wishing to defend the consistency of the classical law of excluded middle, Kolmogorov [1925] gave a nearly complete axiomatization of minimal logic, the intuitionistic propositional logic of $\supset$ and $\neg$.[34] Focusing on the importance of what is now called "stability under double negation," Kolmogorov showed that if a proposition b in this language could be proved using also the classical schema $\neg\neg a \supset a$, then the proposition b^k obtained from b by doubly negating each of its prime components could be proved intuitionistically. This transformation effectively reduced all of classical propositional logic to Kolmogorov's fragment of

[32] Troelstra and van Dalen [1988] used the acronyms **IQC** and **IPC** for different, but equivalent, axiomatizations of intuitionistic predicate and propositional logic respectively.

[33] Classically (but not intuitionistically) $A \vee B$ is equivalent to $\neg(\neg A \wedge \neg B)$, $A \supset B$ is equivalent to $\neg(A \wedge \neg B)$, and $\exists x A(x)$ is equivalent to $\neg\forall x \neg A(x)$.

[34] Like Glivenko [1928], he omitted Heyting's axiom 4.1; cf. [Mancosu *et al.*, 2009].

minimal logic.

Glivenko [1928; 1929] also reduced the consistency of classical propositional logic to that of the intuitionistic theory. If some formula and its negation were both classically provable, then by Glivenko's theorems both the double negation and the negation of that formula would be provable intuitionistically, so intuitionistic propositional logic would be inconsistent. Glivenko's Theorems, as stated in §2.3, do not hold for intuitionistic predicate logic, although Gödel [1933] showed that they do for the intermediate logic obtained by adding as axiom schema the converse of Heyting's theorem 6.78.[35]

Gödel [1933] and Gentzen [1933] independently completed the picture by faithfully translating classical predicate logic into the fragment of intuitionistic predicate logic without the logical symbols $\vee, \exists$, by recursion on the inductive definition of formula (i.e. by "formula induction"). The Gödel translation p^g of a prime formula p may be taken to be $\neg\neg p$, and compound formulas are translated as follows:

$$(a \wedge b)^g \equiv_D a^g \wedge b^g, \quad (a \supset b)^g \equiv_D \neg(a^g \wedge \neg b^g), \quad (a \vee b)^g \equiv_D \neg(\neg a^g \wedge \neg b^g),$$

$$(\neg a)^g \equiv_D \neg a^g, \quad (\forall x\ a)^g \equiv_D \forall x\ a^g \quad \text{and} \quad (\exists x\ a)^g \equiv_D \neg\forall x \neg a^g.$$

It is not hard to show that $a \supset\subset a^g$ is classically provable for every formula a, and that the formula a is provable classically if and only if a^g is provable intuitionistically. The Gentzen translation is similar except that implications are translated by implications.

The negative translation applies also to arithmetic, where prime formulas (equalities) can be taken as their own translations. In intuitionistic arithmetic every formula containing no disjunction and no existential quantifier is *stable* under double negation, in the sense that the formula and its double negation are provably equivalent.

Kleene [1952, p. 497] remarked that van Dantzig [1947] suggested "that it may be possible to interpret practically the whole of classical mathematics within this stable part of the intuitionistic system;" much later P. Krauss [1992, unpublished] carried out part of this plan. In essence, one simply treats the combinations $\forall\neg\neg$ and $\neg\neg\exists$ as classical quantifiers, and the Gödel translation of $\vee$ as classical disjunction, within the intuitionistic language; cf. [Moschovakis, 2003]. The double negation following the universal quantifier is needed to break the uniform dependence of the quantified expression on the variable in question.

The negative translation established the proof-theoretic equivalence of a variety of classical theories with their intuitionistic subtheories. In 1932 Gödel [1932] showed that if classical first-order arithmetic with addition and multiplication is *ω-consistent* (that is, if for no formula $A(x)$ are $\neg\forall x A(x)$, $A(\mathbf{0})$, $A(\mathbf{1})$, $A(\mathbf{2})$, ... all provable), then it is incomplete in the sense that there is a sentence E such

[35] Cf. Troelstra [1973, p. 93], where this converse is called "double negation shift." Troelstra notes that it also justifies the negative translation of the axiom schema of countable choice. However, as Heyting's counterexample (§3.2 above) showed, double negation shift fails intuitionistically when the universal quantifier ranges over choice sequences.

that neither E nor $\neg E$ is provable; and if it is simply consistent then the sentence expressing its consistency is unprovable. These are Gödel's famous First and Second Incompleteness Theorems. In 1936 Church [1936; 1936a] proved that classical predicate logic and classical arithmetic (with $+$ and $\cdot$ and assuming ω-consistency) are undecidable in the sense that there is no recursive procedure for deciding whether or not a given sentence is provable. In the same year Rosser [1936] showed that simple consistency was a strong enough assumption for both Gödel's and Church's theorems. These results hold also for the corresponding intuitionistic subtheories.[36]

Simple consistency of classical (hence also of intuitionistic) predicate logic was established by Hilbert and Ackermann using two-valued truth tables with a one-element domain. After proving his incompleteness theorems, Gödel remarked that it might still be possible to establish the consistency of arithmetic using methods which were finitary in a wider sense, but not formalizable in arithmetic. By rethinking the nature of logical derivation and admitting a rule of inference (the *ω-rule*) with countably infinitely many premises, Gentzen [1935] succeeded in constructing a consistency proof for a system of arithmetic for which consistency and ω-consistency were equivalent by fiat. His method applied directly to intuitionistic as well as to classical ω-arithmetic, by a simple structural restriction which also marked the difference between his formal systems for intuitionistic and classical logic. This brings us to the next topic.

5.2 Constructive Disjunction and Existence

The power of intuitionistic logic lies in the effectiveness or constructiveness of its conclusions. Within reason, negative translations of nonconstructive axioms can be added to an intuitionistic theory without losing this effectiveness; and although the presence of all those double negations may seem clumsy and unnecessary to a classical mathematician, there is a practical value to this approach.

The constructive nature of intuitionistic logic is exemplified by the fact that if $A \vee B$ is provable and closed (i.e., contains no free occurrences of individual variables) then at least one of A, B must be provable. For intuitionistic propositional logic this *disjunction property* [DP] was stated without proof by Gödel [1932a] in 1932 and proved by Gentzen [1935] in 1935 (see the next subsection); Gentzen also proved it for intuitionistic predicate logic. Implicit in Gentzen's work and in Schütte [1950] was the *existence property*:

[EP]: If $\exists x\ A(x)$ is a closed theorem, then $A(t)$ is a theorem for some term t.

for intuitionistic predicate logic.[37] The following strong existence property for intuitionistic arithmetic is due to Kleene [1945, p. 115] and Corollary 2(b) to

[36] See [**?**] for a fuller discussion of Gödel's, Church's and Rosser's theorems.

[37] Since individual variables are the only terms of pure predicate logic, it follows that if $\exists x A(x)$ is a closed theorem of pure intuitionistic predicate logic, so is $\forall x A(x)$. Kleene [1962] observes: "This result was evidently communicated by Gentzen to Heyting in advance of the publication of [Gentzen, 1935] . . . "

Theorem 62 of [Kleene, 1952], with the discussion following) and was reproved by Harrop [1956], extending a partial result by Rasiowa [1955]:

[EPN]: If $\exists x\ A(x)$ is a closed theorem, then $A(\mathbf{n})$ is a theorem for some numeral $\mathbf{n}$.

Here the *numerals* are $0, 0', 0'', 0''', \ldots$.[38]

5.3 Gentzen's Contribution

All the formal systems we have considered so far are *Hilbert-style*, with many axioms and few rules of inference. Proofs and derivations are linear, with each step possibly depending logically on one or more earlier steps. For some uses a *natural deduction* system is preferable, with few axioms and many rules of inference; derivations are presented as trees, clarifying the dependency relations.

Gentzen [1935] introduced a new kind of formal system for propositional and predicate logic, in which *antecedents* (assumptions) as well as *consequents* were listed at each step. Each step in a Gentzen-style derivation was a *sequent* of the form $A_1, \ldots, A_m \rightarrow B_1, \ldots B_s$ where $A_1, \ldots, A_m$ and $B_1, \ldots B_s$ were formulas, and for the intuitionistic system s must always be ≤ 1. The only axiom schema was $C \rightarrow C$. Rules of inference were of two kinds: rules for introducing logical symbols in the antecedent and consequent of a sequent, and structural or bookkeeping rules. We give the rules for introducing negation:[39]

(in consequent): From $A, \Gamma \rightarrow \Theta$ conclude $\Gamma \rightarrow \Theta, \neg A$.

(in antecedent): From $\Gamma \rightarrow \Theta, \mathrm{A}$ conclude $\neg A, \Gamma \rightarrow \Theta$.

Here Γ and Θ are lists of formulas, with Θ empty for the intuitionistic system. The reason for the restriction on s (which follows from the restrictions on Θ) for the intuitionistic system becomes clear if we consider that for $s > 1$ the sequent $A_1, \ldots, A_m \rightarrow B_1 \vee \ldots \vee B_s$ could be derived from $A_1, \ldots, A_m \rightarrow B_1, \ldots, B_s$ and so $\rightarrow A \vee \neg A$ would be provable from $\rightarrow A, \neg A$ which follows from $A \rightarrow A$ by the unrestricted first negation-introduction rule.

The only introduction rule for which the logical form of the hypothesis or hypotheses could not be determined from the logical form of the conclusion was the *cut rule*, which for the intuitionistic system read as follows:

From $\Delta \rightarrow A$ and $B, \Gamma \rightarrow \Theta$ conclude $A \supset B, \Delta, \Gamma \rightarrow \Theta$.

[38] Kleene proved the stronger result that if $\forall x \exists y A(x, y)$ is a closed theorem of **HA**, then there is a recursive function φ such that for every n: if $\varphi(n) = m$ then $A(\mathbf{n}, \mathbf{m})$ is provable.

In arithmetic, disjunction can be defined from existential quantification by $A \vee B \equiv_D \exists x[(x = 0 \supset A) \wedge (\neg(x = 0) \supset B)]$ so the existence property entails the disjunction property. In 1975 H. Friedman [1975] showed that the converse holds also, for any recursively enumerable extension of **HA**.

[39] We follow Kleene's presentation of Gentzen's system G1 in [Kleene, 1952, §77]. For more detail see Kleene [Kleene, 1952, §§77–79] and von Plato's chapter on Gentzen's logic in this volume.

Gentzen proved that every (intuitionistic or classical) derivation in his system could be reduced to a *normal form*, in which the cut rule was not used, and that every formula occurring in any sequent of a normal derivation was a subformula of some formula occurring in the antecedent or consequent of the last sequent. This remarkable result is Gentzen's *Hauptsatz*.

The disjunction property for intuitionistic logic is an immediate consequence. In any Gentzen-style proof of a sentence E, the last sequent would be $\rightarrow E$; and if E was $A \vee B$ then the previous sequent in a normal intuitionistic derivation must have been either $\rightarrow A$ or $\rightarrow B$. To see that the disjunction property fails for classical propositional logic, consider $P \vee \neg P$ where P is prime. Since propositional logic is closed under substitution, the provability of P or of $\neg P$ would yield inconsistency; yet $P \vee \neg P$ is classically provable.

5.4 Admissible Rules and Intermediate Logics

The disjunction and existence properties are special cases of an important structural difference, which has recently been illuminated, between intuitionistic and classical logic. The difference concerns the distinction between derivable and nonderivable admissible rules of inference.

A logical rule is a schema, a shorthand notation for infinitely many particular inferences. If $\mathcal{S}$ is a Hilbert-style formal system closed under substitution and if $A_1, \ldots, A_k, B$, are formulas, the logical rule

$$\text{From } \ A_1, \ldots, A_k \ \text{ conclude } \ B.$$

means that if $\sigma(A_1), \ldots, \sigma(A_k), \sigma(B)$ are obtained by a uniform substitution of particular formulas for the predicate letters occurring in $A_1, \ldots, A_k, B$, and if $\sigma(A_1), \ldots, \sigma(A_k)$ have already been deduced in $\mathcal{S}$ from assumptions Γ, then $\sigma(B)$ can also be deduced in $\mathcal{S}$ from Γ. In predicate logic there may be some restrictions on the free variables. The rule is called *admissible* for $\mathcal{S}$ if, for every substitution σ, if $\sigma(A_1), \ldots, \sigma(A_k)$ are closed theorems of $\mathcal{S}$, then $\sigma(B)$ is also provable in $\mathcal{S}$; the standard notation for this situation is

$$A_1, \ldots, A_k \mid\!\sim_{\mathcal{S}} B\,.$$

The rule is called *derivable* in $\mathcal{S}$ if the conclusion is deducible in $\mathcal{S}$ from the hypotheses, i.e. if $A_1, \ldots, A_k \vdash_{\mathcal{S}} B$.

For pure classical logic there is essentially no difference; every admissible rule of **CPC** is classically derivable, and every nonderivable admissible rule of **CQC** is vacuous in the sense that there is no substitution σ which converts each hypothesis into a closed theorem (cf. [Dzik, 2004]). Intuitionistically the situation is more interesting.

One of the basic tenets of Brouwer's philosophy is that no formal system can adequately represent intuitionistic reasoning, even about mathematics. The purpose of intuitionistic logic is to clarify some general kinds of reasoning which can

always be supported by correct constructions. From the intuitionistic viewpoint, a rule of the form

$$\text{From } A_1, \ldots, A_k \text{ conclude } B.$$

is admissible for a formal system $\mathcal{S}$ if there exists an algorithm which transforms any given k-tuple $\Pi_1, \ldots, \Pi_k$, such that each Π_i is a proof in $\mathcal{S}$ of a sentence of the form $\sigma(A_i)$, into a proof in $\mathcal{S}$ of $\sigma(B)$. The algorithm has no work to do in the case of a substitution σ for which some $\sigma(A_i)$ is unprovable in $\mathcal{S}$, even if $\sigma(A_i)$ can be established by other intuitionistically correct means.

On the other hand, the rule is derivable in $\mathcal{S}$ if and only if there exists a general deduction-schema in $\mathcal{S}$ establishing that B follows in $\mathcal{S}$ from $A_1, \ldots, A_k$. This is evidently a stronger requirement on two counts: it applies to *every* $(k+1)$-tuple of formulas of the form $A_1, \ldots, A_k, B$, and it insists that $\mathcal{S}$ *itself* must provide an algorithm for establishing the correctness of the inference.

The disjunction property, when stated as

$$\text{From } A \vee B \text{ conclude } A \text{ or } B$$

for A, B closed formulas, is an important admissible propositional rule of intuitionistic logic which is completed constructively by an algorithm for producing, from an arbitrarily given proof of any sentence of the form $A \vee B$, either a proof of A or a proof of B. In its presence every other admissible propositional rule can be stated with a unique conclusion.

Two early examples of admissible nonderivable propositional rules of intuitionistic logic are the "independence of premise" rule:

$$\neg A \supset B \vee C \;\vdash\; (\neg A \supset B) \vee (\neg A \supset C).$$

and Mints' rule:

$$(A \supset B) \supset A \vee C \;\vdash\; ((A \supset B) \supset A) \vee ((A \supset B) \supset C).$$

The disjunction rule evidently works together with the other rules to strengthen their conclusions. By generalizing Mints' rule, A. Visser and D. H. J. de Jongh identified a recursively enumerable sequence of successively stronger admissible nonderivable rules of intuitionistic propositional logic and conjectured that they form a *basis* for all such rules in the sense that every admissible rule of propositional logic is derivable from them together with the disjunction property and *modus ponens*. "Visser's rules" are all the rules of the form

$$(A \supset B \vee C) \vee D \;\vdash\; \bigvee_{i=1}^{n} (A \supset E_i) \vee (A \supset B) \vee (A \supset C) \vee D.$$

where n is a positive integer and A is $\bigwedge_{i=1}^{n} (E_i \to F_i)$. Using results of Ghilardi [1999], in 2001 R. Iemhoff [2001] succeeded in proving their conjecture. Since Visser's comprehensive 1999 survey [Visser, 1999] contained a proof that the admissible propositional rules of intuitionistic arithmetic are the same as those of

intuitionistic propositional logic, Iemhoff's result applied also to Heyting arithmetic.

Because its hypothesis as well as its conclusion depends on the provability relation, a rule which is admissible for one formal system may fail for an extension of that system by additional axiom schemas. An *intermediate propositional logic* is any consistent collection of propositional formulas containing all the theorems of intuitionistic logic and closed under *modus ponens* and substitution (of arbitrary formulas for propositional variables). Every intermediate propositional logic is contained in classical propositional logic.

Gödel [1933] introduced an infinite sequence of intermediate logics in order to prove that intuitionistic propositional logic had no finite truth-table interpretation. The nth Gödel logic G_n is axiomatized over **IPC** by the two schemas

$$\bigvee_{i=1}^{n}(\bigwedge_{j=1}^{i-1} A_j \supset A_i) \quad \text{and} \quad (A \supset B) \vee (B \supset A).$$

The Visser rules are derivable in G_n but the disjunction property does not hold. Gabbay and de Jongh [1974] found an infinite sequence of finitely axiomatizable intermediate logics D_n with the disjunction property, such that the first $n+1$ Visser rules (but not the $n+2$nd) are admissible for D_n. Finally Iemhoff [2005] proved that intuitionistic propositional logic is the only intermediate logic with the disjunction property which is closed under all the Visser rules.

The nonderivable admissible predicate rules of intuitionistic predicate logic have not been completely characterized, but they are known to form a complete Π^0_2 set and to include Markov's rule for decidable predicates:

$$\forall x(A(x) \vee \neg A(x)) \wedge \neg\forall x \neg A(x) \quad \mid\!\sim \quad \exists x A(x).$$

and an additional independence of premise rule:

$$\forall x(A(x) \vee \neg A(x)) \wedge (\forall x A(x) \to \exists y B(y)) \quad \mid\!\sim \quad \exists y(\forall x A(x) \to B(y)).$$

where y must not occur free in $\forall x A(x)$. These rules also hold for sentences of intuitionistic arithmetic. Markov's rule and the decidability (hence also the stability under double negation) of quantifier-free arithmetical formulas give the important result [Kreisel, 1958] that classical and intuitionistic arithmetic prove the same formulas of the form $\forall x \exists y A(x,y)$ with $A(x,y)$ quantifier-free, so they have the same provably recursive functions.[40]

An admissible rule with *no* hypotheses is simply a logical theorem. The *propositional logic of* intuitionistic arithmetic consists of those formulas in the language

[40] A Π^0_2 relation is one which can be expressed in $\forall\exists$ form, with a recursive scope. A relation is *complete* Π^0_2 if it is Π^0_2, and every Π^0_2 relation is recursive in it. The standard complete Π^0_2 relation $R(e)$ is "e is the Gödel number of a total recursive function," or equivalently "for every n there is an m which is the Gödel number of a computation of the value of the eth partial recursive function on the argument n."

of propositional logic, all of whose arithmetical substitution instances are provable in Heyting arithmetic. De Jongh [1970] proved that the propositional logic of Heyting arithmetic is **IPC**. To quote Visser [1999], de Jongh actually established "a much stronger result, viz., that the logic of relative interpretations in **HA** is Intuitionistic Predicate Logic ..."[41] His proof was a clever combination of recursive realizability and Kripke semantics, two interpretations of intuitionistic logic and mathematics which will be introduced in the next section.

6 INTERPRETATIONS OF INTUITIONISTIC LOGIC AND ARITHMETIC

After the formalization of intuitionistic logic and Gödel's completeness proof for classical logic, the search for classical interpretations of intuitionistic propositional and predicate logic began. This semantic approach has produced two related interpretations which are sound and complete for **IPC** and **IQC**: algebraic semantics and Beth-Kripke semantics. The constructive viewpoint is probably best represented by Kleene and Nelson's realizability semantics for **HA**, for which it is sound but not complete. We discuss these three interpretations here in historical order.

6.1 Algebraic and Topological Interpretations

According to Troelstra [1991] the first semantics for **IPC**, "due to the work of S. Jaskowski, M. H. Stone, A. Tarski, G. Birkhoff, T. Ogasawara in the years 1936-1940," was an algebraic interpretation in which elements of a Heyting algebra were assigned as truth values to propositions. A Heyting algebra is a lattice with a maximum and a minimum element and an operation $\supset$ satisfying the condition $a \wedge b \leq c$ if and only if $a \leq b \supset c$.

The topological interpretation, developed for **IQC** as well by Mostowski [1948] and by Rasiowa and Sikorski [1954; 1963], uses the algebra of open sets of a topological space T under the partial ordering $\subseteq$, with $G \cup H$, $Int(G \cap H)$ and $Int(H \cup (T \setminus G))$ respectively interpreting $a \vee b$, $a \wedge b$, $a \supset b$ when a is interpreted by G and b by H. The logic then requires that $\neg a$ be interpreted by $Int(T \setminus G)$. For the extension to predicate logic $\exists$ is interpreted by union, and $\forall$ by the interior of the intersection. Soundness guarantees that each closed theorem is always interpreted by T. Completeness asserts that to each unprovable sentence E of **IPC** or **IQC** there corresponds a topological space in which E does not take T as value. The topological interpretation was later extended to (subsystems of) intuitionistic analysis by Scott [1968; 1970], Krol' [1973] and others.

[41] A *relative interpretation* is an interpretation of a theory over a domain which can be represented by a formula $D(x)$ for which the theory proves $\exists x D(x)$. Then $\forall x(D(x) \supset A'(x))$ and $\exists x(D(x) \wedge A'(x))$ interpret $\forall x A(x)$ and $\exists x A(x)$ respectively, if $A'(x)$ interprets $A(x)$.

6.2 Recursive Realizability and Church's Thesis

Kleene's number-realizability interpretation [Kleene, 1945] for intuitionistic arithmetic was inspired by a passage in Hilbert and Bernays [1934], but it can easily be understood as a recursive implementation of the B-H-K explication of the logical connectives and quantifiers. It used the fact that an ordered pair of natural numbers (n, m) can be (primitive) recursively coded by a natural number k in such a way that n and m (call them $(k)_0$ and $(k)_1$, respectively) can be recovered (primitive) recursively from k.[42] It also used the recursively enumerable relation "e is the Gödel number of a partial recursive function which is defined at the argument x with value y," abbreviated $\{e\}(x) \simeq y$.

Kleene associated with each arithmetical sentence E a relation e *realizes* E, defined inductively as follows:

1. e realizes a prime sentence $s = t$ if $s = t$ is true under the intended interpretation.

2. e realizes $A \wedge B$ if $(e)_0$ realizes A and $(e)_1$ realizes B.

3. e realizes $A \vee B$ if $(e)_0 = 0$ and $(e)_1$ realizes A, or $(e)_0 > 0$ and $(e)_1$ realizes B.

4. e realizes $A \supset B$ if whenever f realizes A, then $\{e\}(f)$ is defined and realizes B.

5. e realizes $\neg A$ if e realizes $A \supset 1 = 0$.

6. e realizes $\forall x A(x)$ if for every n, $\{e\}(n)$ is defined and realizes $A(\mathbf{n})$.

7. e realizes $\exists x A(x)$ if $\{(e)_1\}(n)$ is defined and realizes $A(\mathbf{n})$, where $n = (e)_0$.

The definition can be interpreted either constructively or classically. Kleene conjectured, and in [Nelson, 1947] Kleene's student D. Nelson proved, that the universal closure of every axiom of intuitionistic arithmetic is realizable, and that if the universal closure(s) of the premise(s) of a rule of inference are realizable then so is the universal closure of the conclusion. Since $0 = 1$ is not realizable, it follows that every realizable sentence is consistent with **HA**. These results are known as "Nelson's Theorem:"

- Every arithmetical sentence which can be derived in intuitionistic arithmetic from realizable sentences is realizable.

- Every extension of intuitionistic arithmetic by realizable axioms is consistent.

[42] Kleene chose the coding $k = 2^n \cdot 3^m$ so that $(k)_0 = n$ and $(k)_1 = m$ are the first two exponents in the prime factorization of k.

Nelson's theorem established the consistency with **HA** of Church's Thesis [CT_0] in the form

$$\forall x \exists y A(x,y) \supset \exists e \forall x \exists y [T(e,x,y) \wedge \forall z (U(y,z) \supset A(x,z))].$$

where $A(x,y)$ may be any formula containing only x and y free, and $T(e,x,y)$ and $U(y,z)$ are decidable formulas expressing the primitive recursive relations "y is the least gödel number of a computation of the value the partial recursive function with gödel number e takes on the argument x" and "z is the value computed by the computation with gödel number y" respectively. This schema, which is inconsistent with classical arithmetic, holds for the recursive mathematics of A. A. Markov.[43] Markov's Principle [MP] is also realizable, by a proof using Markov's Principle informally. The consistency of **HA** + CT_0 + MP, expressing a large part of Russian recursive mathematics, follows.

Nelson formalized the notion "e realizes E" within a conservative extension $\mathbf{HA}'$ of Heyting arithmetic with a new symbol $\mathbf{r}$, and proved that for all formulas A, B of the language of $\mathbf{HA}'$:

- If A, B are sentences and $A \supset B$ is provable in $\mathbf{HA}'$, so is $\mathbf{r}A \supset \mathbf{r}B$.
- The universal closure C of the formula $(A \supset \mathbf{r}A) \wedge (\mathbf{r}A \supset A)$ is realizable.
- The sentence $\mathbf{r}\,C$ is provable in $\mathbf{HA}'$.

Kleene [1945] discussed, interpreted and extended Nelson's results. An important modification of what is now known as "number-realizability" was the relation e *realizes-*$\vdash E$, obtained by changing the inductive cases for $\vee, \supset$ and $\exists$ as follows:

3′. e realizes-$\vdash A \vee B$ if $(e)_0 = 0$ and $(e)_1$ realizes-$\vdash A$ and $\vdash A$, or $(e)_0 > 0$ and $(e)_1$ realizes-$\vdash B$ and $\vdash B$.

4′. e realizes-$\vdash A \supset B$ if whenever f realizes-$\vdash A$ and $\vdash A$, then $\{e\}(f)$ is defined and realizes-$\vdash B$.

7′. e realizes-$\vdash \exists x A(x)$ if $\{(e)_1\}(n)$ is defined and realizes-$\vdash A(\mathbf{n})$ and $\vdash A(\mathbf{n})$, where $n = (e)_0$.

The analogue of Nelson's theorem holds for realizability-$\vdash$. A more general result for realizability-$(\Gamma \vdash)$ (replacing "$\vdash$" by "$\Gamma \vdash$" in each inductive clause) is

- Every sentence which can be derived in intuitionistic arithmetic from realizable-$\vdash$ sentences Γ is realizable-$(\Gamma \vdash)$.

Realizability-$\vdash$ gives the disjunction and existence properties for **HA**, and the closure of **HA** under *Church's Rule*: If $\forall x \exists y A(x,y)$ is a closed theorem, so is

[43] Church's Thesis was not accepted by Brouwer or Heyting, but this formulation plays a parallel role for recursive mathematics to Brouwer's continuity principle for intuitionistic analysis, to be discussed in a later section.

$\forall x \exists y[T(\mathbf{e}, x, y) \wedge \forall z(U(y, z) \supset A(x, z))]$ for some numeral $\mathbf{e}$. It follows that **HA** can prove the existence only of provably recursive functions.[44]

Kleene initially hoped that realizability might provide a suitable notion of truth in intuitionistic arithmetic. Nelson's work verified that the schema

$$(A \supset \mathbf{r}A) \wedge (\mathbf{r}A \supset A)$$

was consistent with **HA**. In [Troelstra, 1971], by carefully analyzing the inductive definition, Troelstra isolated a schema he called "extended Church's Thesis" [ECT_0] which had the effect over **HA** of identifying truth with realizability. Troelstra argued further that **HA** + ECT_0 + MP accurately represented the practice of Russian recursive mathematics.

By that time Kleene had found that extending his notion smoothly to intuitionistic analysis required the use of functions, not numbers, as realizing objects, and that this altered the class of realizable arithmetical formulas. Many modifications of realizability have since been developed as tools to obtain consistency and independence results for formal systems based on intuitionistic logic. Troelstra's [1998] and van Oosten's [2002], taken together, tell the most important parts of this intriguing story and bring it up to date.

6.3 Kripke Semantics for IPC and IQC

The completeness question for intuitionistic logic is subtle because **IPC** does not admit a finite truth-table interpretation (Gödel [1933]) and every classical model, in the usual sense, of **IQC** is also a model of classical predicate logic. E. W. Beth [1956; 1959] and S. Kripke [1965] each developed a "possible-world" semantics based on allowing decisions about truth or falsity of atomic sentences to be postponed from one moment in time to a later moment, and to be made in different ways along different branches in a partial ordering of possible situations. Each proved a completeness theorem for **IQC**. Kripke's models are simpler than Beth's so we present them first, out of historical order.[45]

A *Kripke model* for intuitionistic propositional logic is a triple $\mathcal{K} = \{K, \leq, \Vdash\}$ where K is an inhabited (i.e. effectively nonempty) set of *nodes* partially ordered by $\leq$, and the *forcing relation* $\Vdash$ is a binary relation between nodes and prime formulas P satisfying the restriction that if $k, k' \in K$ and $k \leq k'$ and $k \Vdash P$ then also $k' \Vdash P$.[46] The forcing relation is extended inductively to all formulas as follows:

1. $k \Vdash A \wedge B$ if $k \Vdash A$ and $k \Vdash B$.

[44] In contrast, **PA** proves $\forall x \exists y[(y = 0 \wedge A(x)) \vee (y = 1 \wedge \neg A(x))]$ for every arithmetical formula $A(x)$.

[45] Kripke observed that his models can easily be transformed into Beth models of a special kind. His completeness proof was based on Beth's, which will be described in a later section.

[46] The "forcing" terminology was not used in this way by Kripke but is now standard. In Kripke's original definition each prime formula P was assigned a truth value T or F at each node k, with the condition that if $k \leq k'$ and P was T at k then also P was T at k'. Here "$k \Vdash A$" corresponds to "A is T at k."

2. $k \Vdash A \vee B$ if $k \Vdash A$ or $k \Vdash B$.

3. $k \Vdash A \supset B$ if, for each $k' \geq k$: if $k' \Vdash A$ then $k' \Vdash B$.

4. $k \Vdash \neg A$ if there is no $k' \geq k$ such that $k' \Vdash A$.

It is straightforward to prove that the forcing relation is *monotone*, consistent, and *sound* for **IPC**:

- For every formula E: If $k \leq k'$ and $k \Vdash E$ then $k' \Vdash E$.
- For no $k \in K$ and no formula E do both $k \Vdash E$ and $k \Vdash \neg E$ hold.
- If $\vdash E$ then $k \Vdash E$ for every $k \in K$.

A Kripke model is a *tree model* if each node k has a unique immediate predecessor in the ordering $\leq$ of K. The tree is *rooted* if there is a (necessarily unique) $k_0 \in K$ such that $k_0 \leq k$ for every $k \in K$. The strong completeness result, from which the decidability of intuitionistic propositional logic follows, is

- To each formula E which is not a theorem of **IPC** there is a finite rooted tree Kripke model whose root k_0 does not force E.[47]

For example, a simple Kripke counterexample to $\neg P \vee \neg\neg P$ is given by $K = \{k_0, k_1, k_2\}$ where $k_0 < k_1$, $k_0 < k_2$ and k_1, k_2 are incomparable, with the forcing relation determined by $k_1 \Vdash P$.

A *Kripke model* for intuitionistic predicate logic is a quadruple $K = \{K, \leq, D, \Vdash\}$ where K is an inhabited set of nodes partially ordered by $\leq$, and for each $k \in K$ the *domain of* k is an inhabited set $D(k)$ of objects with the restriction that if $k, k' \in K$ and $k \leq k'$ then $D(k) \subseteq D(k')$. The forcing relation can be thought of as a binary relation between nodes and prime sentences of an expanded language with an individual constant for each element of $\bigcup_{k \in \mathcal{K}} D(k)$, with the restriction that if $k \in K$ and $k \Vdash P$ where P is prime, then every constant occurring in P is an element of $D(k)$ and if $k' \in K$ with $k \leq k'$ then also $k' \Vdash P$.[48] The forcing relation extends to all sentences of the expanded language using, in addition to the clauses for the propositional connectives, the following:

5. $k \Vdash \forall x A(x)$ if, for each $k' \geq k$ and each $d' \in D(k')$: $k' \Vdash A(d')$.

6. $k \Vdash \exists x A(x)$ if, for some $d \in D(k)$: $k \Vdash A(d)$.

Monotonicity, consistency and soundness extend naturally to the predicate case. The completeness theorem takes the form

[47] For details see [Troelstra and van Dalen, 1988] or [Smorynski, 1973], where the Kripke completeness theorems for **IPC** and **IQC** are presented using the method of Aczel [1968].

[48] Instead of expanding the language, Kripke associated with each node k and each n-place prime formula P a subset of $[D(k)]^n$ on which P was T at k, with the corresponding monotonicity requirement. The two approaches are equivalent classically, and for counterexamples it is convenient to have names in the language for elements of the domain.

- If E is a sentence of the language of predicate logic which is forced at every node of every Kripke model, then E is provable in **IQC**.

The argument actually shows that the assumption that E is unprovable is contradictory; the existence of a proof then follows by informal Markov's Principle. Following Gödel, Kreisel [1962] argued that completeness of intuitionistic predicate logic entails Markov's Principle, but Veldman [1976] found a way to escape this conclusion by modifying Kripke's definition to allow "exploding" nodes at which *all* sentences of the relevant language are forced. The modification involved adding a propositional constant $\perp$ for absurdity to the language and defining $\neg A$ to be $A \supset \perp$, so clause 4 of the definition could be omitted. Because $\perp$ was a new prime sentence, $k \Vdash \perp$ could happen at one or more nodes k of a Veldman model. Because $\vdash \neg\perp$ (because $\vdash \perp \supset \perp$), if $k \Vdash \perp$ then $k \Vdash E$ for every sentence E.

From the classical point of view, a Kripke model is just a Veldman model with all the exploding nodes omitted. But Veldman's intuitionistic completeness proof is a beautiful application of one of the main principles of intuitionistic analysis. Since Beth's completeness proof is best understood in terms of the intuitionistic notion of spread, it is time for us to explore Brouwer's boldest creation.

7 THE SECOND ACT OF INTUITIONISM

From the beginning Brouwer struggled with the problem of the continuum. In his dissertation, while rejecting Kant's claim for an *a priori* intuition of space, Brouwer admitted the need for a separate intuition of the continuum:

> ... the basic intuition of mathematics (and of any intellectual activity) as the substratum, divested of all quality, of any perception of change, a unity of continuity and discreteness, a possibility of thinking together several entities, connected by a 'between', which is never exhausted by the insertion of new entities. ... Having recognized that the intuition of continuity of 'fluidity', is as primitive as that of several things conceived as forming together a unit, the latter being at the basis of every mathematical construction, we are able to state properties of the continuum as a 'matrix of points to be thought of as a whole'. [Brouwer, 1907, p. 17]

Then, after showing how to construct from the integers the rational numbers with their dense linear ordering (of order type η), he affirmed a form of completeness with the following justification:

> The mathematical intuition is unable to create other than denumerable sets of individuals. But it is able, after having created a scale of ordertype η, to superimpose upon it a *continuum as a whole*, which afterwards can be taken conversely as a measurable continuum, which is the matrix of the points on the scale. [Brouwer, 1907, p. 45].

The problem was to reconcile this independent intuition of the continuum with the fact that the only real numbers Brouwer then considered to be legitimate mathematical entities were the definable real numbers. He believed that only finite or denumerably infinite sets could be constructed, and he accepted Cantor's diagonal argument for the nondenumerability of the continuum. Moreover, every denumerable set must have measure zero, while the continuum should have positive measure.[49] This conflict led the young Brouwer to the theoretical notion of a "denumerably unfinished set" to describe such aggregates as Cantor's second number class, "the totality of definable points on the continuum, and a fortiori the totality of all possible mathematical systems" [Brouwer, 1907, p. 82]. Such sets, which would have to consist of more than denumerably many individual elements, could not exist mathematically and so could not be put in one-to-one correspondence with the continuum.

From the foundational viewpoint Brouwer was not satisfied with this *ad hoc* solution. In [Brouwer, 1918] and [Brouwer, 1919] he finally succeeded in developing an intuitionistic structure for the continuum, or more precisely its underlying *Baire space* consisting of the infinitely proceeding sequences of natural numbers with the initial-segment topology. His construction could be interpreted classically or intuitionistically; however, by considering the intuitionistically acceptable ways of proving properties of the entire structure, Brouwer was led to two new principles one of which contradicted the natural classical interpretation. With this bold step he established intuitionistic analysis as an alternative to classical analysis, not just a constructive part thereof.

7.1 Species, Spreads and Infinitely Proceeding Sequences

In 1918 and 1919 Brouwer published two fundamental papers with the title "Foundation of set theory independent of the logical principle of the excluded middle" (*Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten*). The first ([Brouwer, 1918]) had the subtitle "General set theory" (*Allgemeine Mengenlehre*) and the second [Brouwer, 1919] "Theory of sets of points" (*Theorie der Punktmengen*). He began by distinguishing between the notions *Species* and *Menge*. A structured set (*Menge*) was a mathematical entity of a definite kind, while a species of first order (*Spezies erster Ordnung*) was "a property which only a mathematical entity can possess, in which case it [the entity] is called an *element of the species of first order* ..." [Brouwer, 1918, p. 4]. A species of second order was a property which could be possessed by species of first order, and so on.

Brouwer explicitly rejected the comprehension axiom of classical set theory, according to which each property determined a set (or a subset of a given set). Instead, he accepted sets of *choice sequences*, infinitely proceeding sequences of

[49] Sixty years later Errett Bishop [1967] succeeded in developing a constructive measure theory on a continuum consisting only of constructive real numbers, making implicit use of the fact that these numbers cannot be enumerated constructively.

mathematical entities previously constructed, where each successive element was allowed to be chosen more or less freely, restricted only by the finitely many choices already made and by the law determining the set. The English term eventually chosen by Brouwer [1954a] for this kind of structured set was "spread." Here is the definition he gave in [Brouwer, 1918]:[50]

> A *spread* is a *law* on the basis of which, when again and again an arbitrary integer (*Ziffernkomplex*) is chosen from the sequence ζ [of the positive integers $1, 2, 3, \ldots$], each of these choices produces either a particular symbol (*Zeichen*), or nothing, or else leads to the termination (*Hemmung*) of the process and the definitive annihilation of its results, where for each n after an unterminated sequence of $n-1$ choices at least one number can be specified which, if it is chosen as the nth number, does *not* lead to the termination of the process. Every sequence (which therefore in general is not representable as finished) of symbols chosen in this way from the spread is called an *element of the spread*. We will also denote the common manner of generation of the elements of a spread M briefly by *the spread* M.[51]

The spread law is twofold, consisting of a *choice law* determining which integers are available to be chosen at each stage of the process, and a *correlation law* determining which symbol is attached to each permitted finite sequence of choices. Brouwer left open the interpretation of "law" in this definition, and it is consistent to assume that every choice sequence determines a spread of which it is the only element.[52] Certainly a spread law may be such that after a particular sequence of $n-1$ permitted choices, every subsequent choice is determined in advance and the process never terminates.

In his popular but detailed exposition "Intuitionism: An Introduction" [Heyting, 1956] Heyting simplified the definition of spread by omitting the possibility of termination (which could be indicated instead by assigning a special symbol and requiring all future choices to be 1, for example), and wrote:

> ... by an *infinitely proceeding sequence* (abbreviated: ips) we mean exactly what the words express, i.e. a sequence that can be continued ad infinitum. The question how the components of the sequence are successively determined, whether by a law, by free choices, by throwing a die, or by some other means, is entirely irrelevant.

Infinitely proceeding sequences whose components were completely determined by a (finite) definition were called by Brouwer "fundamental sequences" or "sharp

[50] The translation from the German is mine.

[51] [Brouwer, 1918, p. 4]. In his introduction to Part I of [Mancosu and van Stigt, 1998] van Stigt points out that the idea of a structured set was already present in Brouwer's 1914 review of a book by Schoenflies and Hahn. The same volume contains van Stigt's English translation of Brouwer's 1921 Dutch summary of [Brouwer, 1918] and [Brouwer, 1919], including a paragraph similar to this.

[52] This relative interpretation is implicit in Kleene and Vesley's [1965].

arrows." Other authors have called them "lawlike." Kreisel [1968] introduced a complementary notion of "lawless" or "absolutely free" choice sequence, obeying the second-order restriction that no restriction would ever be imposed on the choices allowed.[53]

A *fan* or *finitary spread* is one with the property that after each sequence $a_1, \ldots, a_{n-1}$ of successive choices permitted by the choice law, only finitely many integers (how many and which ones depending on the choices already made) are available to be chosen as the value of a_n. A fan is the intuitionistic analogue of a compact set in classical mathematics. At the other extreme is the *universal spread*, in which every natural number is available to be chosen at each stage of the construction, and the symbol correlated after each successive choice is the numeral for the number just chosen.

Brouwer was especially interested in the spread of *real number generators*, each representing (in one version) an infinitely proceeding sequence of nested closed intervals with rational endpoints and lengths converging toward zero, or (in another) a Cauchy sequence $\{r_n\}$ of rational numbers.[54] In particular, a *canonical* real number generator can be defined (cf. [Heyting, 1956, p. 41]) as an ips $\{x_n 2^{-n}\}$, where the choice law determining which sequences $\{x_n\}$ of integers may be chosen satisfies

$$| x_n 2^{-n} - x_{n+1} 2^{-n-1} | \leq 2^{-n-1}$$

or equivalently $| 2x_n - x_{n+1} | \leq 1$. The finite initial segments of such a sequence would give better and better rational approximations to the real number in the process of generation, and this was the only information one could expect to have about an arbitrary real number generator. Thus the general question whether any two real number generators coincide, or converge to the same limit, is intuitionistically undecidable.[55]

By asserting that spreads and their elements were legitimate mathematical entities, Brouwer accepted infinitely proceeding sequences of natural numbers, even though "most" such sequences had no possibility of ever being completely defined. By accepting species of every finite order he made it possible to compare the intuitionistic theory of properties with classical abstract set theory. This broadening of the concept of a legitimate mathematical entity, with the new insights it

[53] In [Kreisel and Troelstra, 1970] Kreisel and Troelstra developed a formal theory of choice sequences based on these two concepts.

[54] $\{r_n\}$ is *Cauchy* if for every positive integer k there is an n such that for all p: $| r_n - r_{n+p} | < 1/k$.

[55] Two real number generators $r \equiv \{r_n\}$ and $s \equiv \{s_n\}$ *coincide* (notation: $r = s$) if for every positive integer k there is an n such that for all p: $|r_{n+p} - s_{n+p}| < 1/k$. If, on the other hand, there is a k so that for all p: $r_{n+p} > s_{n+p} + 1/k$, then $r > s$ (and $s < r$). The classical trichotomy law $r < s \vee r = s \vee r > s$ fails intuitionistically, but the following useful substitute holds: If $r < s$ then $r < t \vee t < s$ for every real number generator t (cf. [Heyting, 1956, p. 25]). *Apartness* of real number generators is a positive notion; precisely, r is apart from s if there exist k and n such that for all p: $|r_{n+p} - s_{n+p}| > 1/k$. It follows that r is apart from s just in case $r < s \vee r > s$, so if r is *not* apart from s then r and s coincide. Some of these results appeared in [Brouwer, 1930], the published version of a lecture Brouwer gave in Vienna in March 1928, with Gödel in the audience.

engendered, was called by Brouwer "the second act of intuitionism."

7.2 Brouwer's Fan Theorem, Continuity Principle, and Bar Theorem

Brouwer was not an easy expositor of his own ideas and he published mostly in Dutch or German. The first edition [Heyting, 1956] of Heyting's delightful "Intuitionism: An Introduction" appeared in 1956, clarifying the intuitionistic approach to real numbers. After describing in detail the arithmetic of real number generators, Heyting proved that every real number generator (abbreviated r.n.g.) coincides with a canonical r.n.g. (abbreviated c.r.n.g.). The choice law for the spread of the c.r.n.g. allows any integer to be chosen as x_1 and from then on requires $x_{n+1} \in \{\, 2x_n - 1,\, 2x_n,\, 2x_n + 1 \,\}$, so except for the first choice it is finitary.

If $x = \{x_n 2^{-n}\}$ and $y = \{y_n 2^{-n}\}$ are c.r.n.g., then $x < y$ if and only if for some n: $x_n + 2 < y_n$, and then $y_{n+p}2^{-(n+p)} - x_{n+p}2^{-(n+p)} > 2^{-n}$ for all p. If a and b are c.r.n.g., the *closed interval* $[\, a, b \,]$ consists of all c.r.n.g. x such that $\neg(x < a \wedge x < b)$ and $\neg(x > a \wedge x > b)$. The basic representation theorem, stated and proved in [Heyting, 1956, p. 42], where it is attributed to Brouwer [1919; 1924], is then

> Theorem [1]. Every closed interval of the continuum coincides with a finitary spread [of canonical real number generators].[56]

Next Heyting stated and proved "Brouwer's Fan Theorem" in a form which is not strictly true classically but which illustrates the intuitionistic concept of a total function:[57]

> Theorem [2]. If an integer-valued function $\varphi(\delta)$ is defined for every element δ of a finitary spread S, then a natural number $[N]$ can be computed from the definition of φ, such that $\varphi(\delta)$ is determined by the first N components of δ; that is, if δ_1 and δ_2 are such elements of S that the first N components of δ_1 are equal to the first N components of δ_2, then $\varphi(\delta_1) = \varphi(\delta_2)$.

As references Heyting cited Brouwer's [1923b; 1924; 1927; 1952] and [1954]. In the proof he worked with the underlying fan K of infinitely proceeding sequences d of integers allowed by the choice law of S, and with the function $f(d) = \varphi(\delta)$ where δ is the element of S associated with d by the correlation law.

First he observed that "As $f(d)$ must be calculable, its value must be determined by a finite number of the components of d; that is to say, by a sequence $a(d)$ in F"

[56] There is a subtlety here, as each rational number of the form $m2^{-n}$ is the limit of *three distinct* c.r.n.g., including one approaching it from the left and one from the right; and these approximating sequences coincide.

[57] In mathematical terms, these two theorems together assert that every function defined everywhere on a closed bounded interval is uniformly continuous. Classically it is true that every *continuous* total function on a closed bounded interval is uniformly continuous. Brouwer's intuition led him to the additional assumption that every (calculable) total function is continuous.

(where F is the species of finite sequences of integers which are permitted by the choice law). This is an instance of Brouwer's continuity principle (discussed below). Then Heyting continued: "Let C be the species of the $a(d)$ which correspond in this way to the elements of K; then every element of K has a segment $a(d)$ in C. Also, if b is a sequence in F, every K-continuation of b has a segment in C; we shall express this property by saying that b is *K-barred* by C."

At this point a classical mathematician would invoke König's Lemma, which says in this context that if a fan has unbarred finite branches of arbitrary length then it has an infinite unbarred branch, to conclude by *reductio ad absurdum* that there is an N such that every b in F has a segment in C of length $\leq N$. But the use of proof by contradiction to establish a positive conclusion (here, the existence of N) is not allowed intuitionistically. Instead, Heyting invoked Brouwer's "Bar Theorem" (explained below) to conclude that such an N exists, completing the proof.

The heart of Brouwer's reconstruction of analysis was contained in his continuity principle and his principle of bar induction. The continuity principle depended on a strictly finitary interpretation of the concept of a total function on a spread, according to which the value such a function takes on an arbitrary element δ of the spread must be determined on the basis of a finite initial segment of the underlying i.p.s. d to which δ is correlated. This conflicted with the classical existence of discontinuous functions, for example the function defined on all infinite sequences of 0s and 1s which takes the value 0 on the identically zero sequence, and 1 on every sequence which is not identically zero, since any finite sequence of 0s can be followed by a 1 and then continued indefinitely. In short, the continuity principle provided a proof of

$$\neg\forall\delta(\forall x\delta(x) = 0 \vee \neg\forall x\delta(x) = 0),$$

giving a strong counterexample to the classical law of excluded middle.

The principle of bar induction (which Brouwer called the "Bar Theorem") for the universal spread of all choice sequences of natural numbers can be stated as follows. If to each i.p.s. α of natural numbers a natural number n_α can be calculated such that the finite sequence $\langle\alpha(1), \ldots, \alpha(n_\alpha)\rangle$ belongs to the species C (called a *bar* for the universal spread), and if each element of C is also an element of the species A, where A has the property that if $\langle a_1, \ldots, a_{n-1}, k\rangle \in A$ for every k then also $\langle a_1, \ldots, a_{n-1}\rangle \in A$, then the empty sequence $\langle\ \rangle$ belongs to A. The principle follows for every spread, in particular for every fan. Brouwer gave a proof of the principle based on an analysis of the types of inference which might be used to prove properties A of choice sequences. In the first edition of [Heyting, 1956] Heyting repeated this argument, but in the second edition (published in 1966) he accepted Kleene's conclusion that the "Bar Theorem" was actually a new assumption and that the bar had to be detachable (see below).

7.3 Kleene and Vesley's Formalism for Intuitionistic Mathematics

Even before Heyting's [1956] appeared, Kleene had been working on an axiomatization of intuitionistic analysis with the goal of making Brouwer's reasoning comprehensible and accessible, especially to classical mathematicians. In 1965 Kleene and Vesley published "The Foundations of Intuitionistic Mathematics, Especially in Relation to Recursive Function Theory" containing a complete axiomatization of the theory of infinitely proceeding sequences including real number generators. The formal system **FIM** was based on a two-sorted language, with variables $x, y, \ldots$ for numbers and $\alpha, \beta, \ldots$ for choice sequences. Its mathematical axioms included those of Heyting arithmetic (with additional primitive recursive functions), an axiom schema of countable choice (or alternatively a comprehension schema asserting the existence of definable functions), a strong axiom schema of continuous choice which Kleene called "Brouwer's Principle for Functions," and an axiom schema of bar induction which Kleene showed to be independent of the other axioms.

Vesley worked out the application to real number generators. Using functions rather than numbers as realizing objects, Kleene succeeded in giving a consistency proof for **FIM** relative to its classically correct subsystem **B** (without Brouwer's Principle). Although **FIM** is inconsistent with the classical law of excluded middle, Kleene's realizability interpretation and Vesley's thorough development of the theory of real number generators justified Brouwerian analysis as a coherent alternative to classical analysis.

Brouwer's Principle made it possible to prove strong undecidability results within intuitionistic analysis, obviating the need for weak counterexamples such as the one Heyting used to show the unprovability, in his predicate logic, of $(x) \neg \neg a \supset \neg \neg (x)\, a$. If $A(\alpha, x)$ is the formula $\alpha(x) = 0$ then the following are formal theorems of **FIM** which contradict classical logic:

- $\vdash \neg\forall\alpha[\forall x A(\alpha, x) \vee \neg\forall x A(\alpha, x)]$ (*27.17).
- $\vdash \neg\forall\alpha[\neg\exists x \neg A(\alpha, x) \vee \neg\neg\exists x \neg A(\alpha, x)]$ (*27.18).
- $\vdash \neg\forall\alpha[\forall x(A(\alpha, x) \vee \neg\forall x A(\alpha, x)) \supset \forall x A(\alpha, x) \vee \neg\forall x A(\alpha, x)]$ (*27.19).

If $B(\alpha)$ is $\forall x A(\alpha, x) \vee \neg\forall x A(\alpha, x)$ then $\vdash \forall\alpha\neg\neg B(\alpha)$ (cf. formal theorem 4.8 in Heyting's propositional logic) and so $\vdash \neg[\forall\alpha\neg\neg B(\alpha) \supset \neg\neg\forall\alpha B(\alpha)]$ by *27.17, giving a strong counterexample to double negation shift in place of Heyting's weak one. Here, of course, we are implicitly using the consistency of **FIM**.

Realizability and its variants provide relative consistency and independence proofs for a variety of principles which have been proposed to extend intuitionistic analysis. A strong version $\forall\alpha[\neg\forall x \neg A(\alpha, x) \supset \exists x A(\alpha, x)]$ of Markov's Principle, while realizable classically and hence consistent with **FIM**, is unprovable (Kleene [1965, p. 131], using a modification of function-realizability inspired by Kreisel [1959]). A weak form $\forall\alpha\neg\neg GR(\alpha)$ of Church's Thesis is also consistent with and independent of **FIM**, where $GR(\alpha)$ is a formula of the form

$\exists e \forall x \exists y[T(e,x,y) \wedge \forall z(U(y,z) \supset \alpha(x) = z)]$ expressing "α is general recursive" (J. R. Moschovakis [1971], using a modified function-realizability interpretation inspired by Vesley [1970]). The strong form of Church's Thesis conflicts with Brouwer's Principle, and in the presence of Markov's Principle the weak form conflicts with the Bar Theorem (cf. [Moschovakis, 2003]).

Troelstra [1973] proposed an extension GC (for "generalized continuity") of Brouwer's Principle, and used it to characterize function-realizability just as he had used ECT_0 to axiomatize number-realizability. On the other hand, beginning with the observation that **FIM** is consistent with classical arithmetic, Kleene [1965] surveyed the possibilities of extending intuitionistic analysis consistently in the classical direction. For other examples see [Troelstra, 1998] and [van Oosten, 2002].

7.4 Heyting's Arithmetic of Species and HA$^\omega$

In [Heyting, 1956, p. 37] Heyting repeated Brouwer's definition from [Brouwer, 1918] ("Definition 1") of a species as "a property which mathematical entities can be supposed to possess," and added:

> Definition 2. After a species $\mathcal{S}$ has been defined, any mathematical entity which has been or might have been defined before $\mathcal{S}$ and which satisfies the condition $\mathcal{S}$, is a *member* of the species $\mathcal{S}$.

The "or might have been" allows for the possibility of infinite species of natural numbers or choice sequences. Heyting gives examples: "the property [of a real number generator] of coinciding with a given number-generator is a species," and "The components of an ips ξ of natural numbers form a species ...." These are species of type 0, while the continuum (consisting of species of type 0, as in the first example) is a species of type 1.

If S and T are species such that every element of T is also an element of S, then T is a *subspecies* of S ($T \subseteq S$), and $S - T$ is the subspecies of those elements of S which cannot belong to T. Two species S and T are *equal* if $S \subseteq T$ and $T \subseteq S$ (so equality of species is extensional). "If $T \subseteq S$ then $S' = T \cup (S - T)$ is not always identical with S, for S' contains only those elements of S for which it can be decided whether they belong to T or not." However, S' is *congruent* to S (in the sense that neither species can contain an element which does not belong to the other). If $S' = S$ then T is a *detachable* subspecies of S. There are many detachable subspecies of the natural numbers, but "the continuum has no other detachable subspecies than itself and the null species."

Troelstra [1973] expanded on Heyting's presentation in [Heyting, 1956] of the intuitionistic theory of species, providing a formal system $\mathbf{HAS}_0$ extending **HA** with variables for numbers and species of numbers, formulating axioms EXT of extensionality and ACA of arithmetical comprehension, and proving that $\mathbf{HAS}_0$ + ACA + EXT is conservative over **HA**. In [Troelstra, 1973a] he stated and proved

the *uniformity principle* [UP]:

$$\forall X\, \exists n\, A(X,n) \supset \exists n\, \forall X\, A(X,n),$$

explicitly contradicting classical second-order logic. A similar principle [UP_1] holds for the intuitionistic theory of species of choice sequences $\mathbf{HAS}_1$, where the lowest type of variable ranges over infinitely proceeding sequences of natural numbers. The uniformity principle implies that the only detachable subspecies of an arbitrary species of type 1 are the species itself and the null species, generalizing Heyting's comment quoted above.

Another extension of **HA** explored by Troelstra and others is intuitionistic arithmetic of arbitrary finite type $\mathbf{HA}^\omega$ (called $\mathbf{N}$ - $\mathbf{HA}^\omega$ in [Troelstra, 1973]), with variables of type 0 (over natural numbers) and for each pair σ, τ of types also variables of type (σ, τ) (over functions from objects of type σ to objects of type τ). Equality between objects of the same type may be taken to be intensional, hence decidable (in $\mathbf{I}$ - $\mathbf{HA}^\omega$, where $=$ expresses identity) or extensional (in $\mathbf{E}$ - $\mathbf{HA}^\omega$) as in Gödel's Dialectica interpretation and the theory of choice sequences. The axioms include those of first-order Heyting arithmetic with induction extended to arbitrary formulas of the language, and schemas defining the function constants of projection, application and primitive recursion at all appropriate types.[58]

7.5 Beth Models and Veldman's Completeness Proof for IQC

A Beth model for **IQC** is a quintuple $\mathcal{M} = \{B, \leq, k_0, D \Vdash\}$ where $(B, \leq)$ is a finitely branching tree with root k_0, the domain function D maps each $k \in B$ to a set of positive integers where $D(k) \subseteq D(k')$ if $k \leq k'$, and the forcing relation for prime sentences P (of the language of **IQC** enhanced by constants for the positive integers) satisfies the condition that $k \Vdash P$ *if and only if* every branch through k eventually encounters a node k' such that $k' \Vdash P$. Beth forcing extends inductively to compound formulas in the same way as Kripke forcing except in the cases of disjunction and existence, which are treated as follows:

2. $k \Vdash A \vee B$ if every branch through k eventually hits a node k' such that $k' \Vdash A$ or $k' \Vdash B$.

6. $k \Vdash \exists x A(x)$ if every branch through k eventually hits a node k' such that, for some $d \in D(k')$: $k' \Vdash A(d)$.

In effect, $k \Vdash A \vee B$ if the subtree through k is barred by the disjoint union of two species C and D, where each node in C forces A and each node in D forces B; and similarly for $\exists$. Beth [Beth, 1959, p. 447] gave his definition in terms of validity on subtrees, so e.g. $A \vee B$ is valid on $\mathcal{M}$ if $\mathcal{M}$ is the union of finitely

[58] See the discussion in Beeson [1985], a rich source of information about formalisms proposed before 1985 for theories based on intuitionistic logic, including intuitionistic set theory and Bishop's constructive mathematics. Troelstra's [1977] discusses choice sequences informally from a variety of perspectives; cf. [Kreisel and Troelstra, 1970].

many (compatible) subtrees on each of which either A or B is valid. Fan induction (bar induction on a finitary spread), or classical König's Lemma, establishes the equivalence of his definition with the one given here.[59]

Monotonicity and soundness are proved by the usual inductive arguments. The completeness theorem asserts that every sentence E which is unprovable in **IQC** has a Beth countermodel, i.e. a Beth model whose root does not force E. The proof consists of a systematic search for a proof of E, using a diagram called a *Beth tableau* to guide the construction of a finitely branching tree with sets of formulas attached to the nodes. If the tableau is closed, giving a proof of E, then every branch of the tree is finite; if not, then the resulting tree is a Beth countermodel to E.

Citing Dyson and Kreisel [1961], Veldman [1976] observed that Beth's construction was inadequate at least from the intuitionistic viewpoint, although Beth's ideas were used in Kripke's classical completeness proof. By modifying Kripke semantics to allow exploding nodes (see §6.3 for the definition), Veldman succeeded in giving an intuitionistic proof of "the existence of a universal modified Kripke-model, i.e. a model such that every sentence, true in the model, is derivable in intuitionistic predicate calculus" while every intuitionistically provable sentence is true in the model. A discussion of his method seems a suitable ending for this historical exposition of the logic of Brouwer and Heyting.

Aczel [1968] used the notion of saturation to provide a perspicuous classical proof of the Kripke completeness of **IQC**. A set Γ of sentences of the language $\mathcal{L}[C]$, the pure predicate language (with $\bot$ instead of $\neg$) expanded by a set C of individual constants, is *C-saturated* if it satisfies the following conditions:

(i) Γ has at least one element.

(ii) If $\Gamma \vdash A \vee B$ where $A \vee B \in \mathcal{L}[C]$, then $A \in \Gamma$ or $B \in \Gamma$.

(iii) If $\Gamma \vdash \exists x A(x)$ where $\exists x A(x) \in \mathcal{L}[C]$, then $A(c) \in \Gamma$ for some constant $c \in C$.

Condition (ii) guarantees that a C-saturated set is closed under deducibility, since if $\Gamma \vdash A$ then also $\Gamma \vdash A \vee A$. For Veldman's proof we do *not* require that Γ be consistent.

To construct the universal Veldman model we need a denumerably infinite set $C = \{c_{i,k}\}$ of distinct individual constants, and an enumeration $A_0, A_1, A_2, \ldots$ without repetition of the sentences of $\mathcal{L}[C]$ such that each A_n contains only constants $c_{i,k}$ with $i, k < n$, or equivalently with $j(i,k) < j(0,n)$ where j is the surjective pairing function obtained from the enumeration

$$\langle 0,0\rangle; \langle 0,1\rangle, \langle 1,0\rangle, \langle 1,1\rangle; \langle 0,2\rangle, \langle 1,2\rangle, \langle 2,0\rangle, \langle 2,1\rangle, \langle 2,2\rangle; \langle 0,3\rangle, \ldots$$

The simplest solution is to take $c_{i,k} = j(i,k)$, so $C = \{0,1,2,\ldots\}$. We need the corresponding projection functions j_0, j_1 so $j_0(j(i,k)) = i$, $j_1(j(i,k)) = k$ and

[59] In their detailed presentation, Troelstra and van Dalen [Troelstra and van Dalen, 1988, p. 677f.] observe that Beth used a constant domain function while van Dalen [1984] allowed expanding domains.

$j(j_0(h), j_1(h)) = h$. We need an enumeration $d_0, d_1, \ldots$ of all deductions in **IQC**, of sentences of $\mathcal{L}[C]$ from finite sets of sentences of $\mathcal{L}[C]$. We need a coding of finite sequences of natural numbers (for example the primitive recursive coding from [Kleene and Vesley, 1965]) and a concatenation operation $*$ so that if w, v code $\langle a_0, \ldots, a_{n-1}\rangle$ and $\langle a_n, \ldots, a_{n+m-1}\rangle$ respectively then $w * v$ is a code for $\langle a_0, \ldots, a_{n+m-1}\rangle$. If α is a function from numbers to numbers then $\overline{\alpha}(n)$ codes $\langle \alpha(0), \ldots, \alpha(n-1)\rangle$. All these ingredients of the proof are effective.

The universal model is based on a spread with spread-law Σ and correlation law $\langle G, \Gamma\rangle$ defined step by step so that for each sequence code w:

(i) $\Sigma(w) \in \{0, 1\}$, and w is permitted if and only if $\Sigma(w) = 0$.

(ii) If $\Sigma(w) = 0$ then $G(w) \subseteq C$ and $\Gamma(w)$ is a finite set of sentences of $\mathcal{L}[G(w)]$ and there is at least one q such that $\Sigma(w * \langle q\rangle) = 0$.

(iii) If $\Sigma(w * \langle q\rangle) = 0$ then one of the following happens:

(a) $G(w) \subsetneq G(w * \langle q\rangle)$ and $\Gamma(w) = \Gamma(w * \langle q\rangle)$,

(b) $G(w) = G(w * \langle q\rangle)$ and $\Gamma(w) \subsetneq \Gamma(w * \langle q\rangle)$.

(c) $G(w) = G(w * \langle q\rangle)$ and $\Gamma(w) = \Gamma(w * \langle q\rangle)$.

The definition begins with $\Sigma(\langle\ \rangle) = 0$, $G(\langle\ \rangle) = \{c_{(0,0)}, c_{(0,1)}, c_{(0,2)}, \ldots\}$, and $\Gamma(\langle\ \rangle) = \emptyset$ and the construction is arranged so that $\Sigma(w * \langle q\rangle)$ depends on whether the length of w is of the form $3k$, $3k+1$ or $3k+2$. In the $3k$ case $q = 0$ is allowed, and (iii)(c) holds, if and only if it is *not* the case that ($A_{j_1(k)}$ is a sentence of $\mathcal{L}[G(w)]$ and d_i proves $\Gamma(w) \vdash A_{j_1(k)}$ for some $i < 3k$); while $q = 1$ is allowed if and only if $A_{j_1(k)}$ is a sentence of $\mathcal{L}[G(w)]$ (in which case $A_{j_1(k)}$ is put into $\Gamma(w * \langle 1\rangle)$). In the $3k+1$ case every q is allowed, and the finitely many distinct new constants $c_{(k+1,0)}, \ldots, c_{(k+1,q)}$ from C are put into $G(w * \langle q\rangle)$ so (iii)(a) holds; this ensures the availability of new constants for witnesses. In the $3k+2$ case we look back to the $3k$ step and check whether $A_{j_1(k)}$ was added to Γ then, and if so whether $A_{j_1(k)}$ was a disjunction or an existential sentence or neither, and decide accordingly whether to allow only $q = 0$ with (iii)(c), or two or infinitely many $q \geq 1$ with (iii)(b) enforcing the disjunction or existence property respectively.

Then the universal modified Kripke model $\mathcal{V} = \{V, \leq, D, \Vdash\}$ is defined by setting
$V = \{\alpha : \forall x \Sigma(\overline{\alpha}(x)) = 0\}$ where

(iv) For each $\alpha \in V$: $\Gamma_\alpha = \bigcup_x \Gamma(\overline{\alpha}(x))$ and $G_\alpha = \bigcup_x G(\overline{\alpha}(x))$ and the domain $D(\alpha)$ is the set of all constants occurring in Γ_α (so $D(\alpha) \subseteq G_\alpha$).

(v) If $\alpha, \beta \in V$ then $\alpha \leq \beta$ if and only if $\Gamma_\alpha \subseteq \Gamma_\beta$.

(vi) For each $\alpha \in V$ and each prime sentence P of $\mathcal{L}[D(\alpha)]$ (including $\bot$): $\alpha \Vdash P$ if and only if $P \in \Gamma_\alpha$ or $\bot \in \Gamma_\alpha$.

The construction was designed to have the following properties, for each $\alpha \in V$ and each n and m:

(vii) Γ_α is $D(\alpha)$-saturated.

(viii) If A_n, A_m are sentences of $\mathcal{L}[D(\alpha)]$ then $A_n \supset A_m \in \Gamma_\alpha$ if and only if for each $\beta \geq \alpha$: if $A_n \in \Gamma_\beta$ then also $A_m \in \Gamma_\beta$.

(ix) If A_n is a sentence of $\mathcal{L}[D(\alpha)]$ of the form $\forall x B(x)$, then $A_n \in \Gamma_\alpha$ if and only if for all $\beta \geq \alpha$: $B(a) \in \Gamma_\beta$ for all $a \in D(\beta)$.

The "only if" directions of (viii) and (ix) are immediate from (vii). Proving the "if" direction of each requires the construction of an appropriate subfan of the spread Σ. The subfan for (viii) is chosen so that each branch β satisfies

(a′) $\beta(3j(0,n)) = 1$ (so $A_n \in \Gamma_\beta$).

(b′) $\beta(3k) \geq \alpha(3k)$ for all k (so if $A_{j_1(k)} \in \Gamma(\overline{\alpha}(3k+1))$ then also $A_{j_1(k)} \in \Gamma(\overline{\beta}(3k+1))$, and so $\Gamma_\alpha \subseteq \Gamma_\beta$).

(c′) $\beta(1) = max(\alpha(1), m, n) + 2$, and $\beta(3k+1) = max(\alpha(3k+1), \beta(3k-2), m, n) + 2$ for all $k \geq 1$ (so $G(\overline{\beta}(3k+2))$ always contains two fresh constants).

(d′) $\beta(3k+2) = \alpha(3k+2)$ except in the case that $\alpha(3k) = 0$ and $\beta(3k) = 1$ and $A_{j_1(k)}$ is a disjunctive or existential sentence; then $\beta(3k+2) \in \{1, 2\}$ (if $A_{j_1(k)}$ is a disjunction) or $\beta(3k+2) \in \{c, d\}$ where c, d are the two fresh constants (if $A_{j_1(k)}$ is existential).

Now suppose A_n, A_m are such that for each $\beta \in V$ with $\alpha \leq \beta$: If $A_n \in \Gamma_\beta$ then $A_m \in \Gamma_\beta$. Then by (a′) and (b′) with the Fan Theorem there is an x_0 such that $A_m \in \Gamma(\overline{\beta}(x_0))$ for every β in the subfan. By fan induction using (b′) and (d′): if β is in the subfan and $x \leq x_0$ then $\Gamma(\overline{\beta}(x)), \Gamma_\alpha, A_n \vdash A_m$. Taking $x = 0$ and using the Deduction Theorem, finally $\Gamma_\alpha \vdash A_n \supset A_m$.

The proof of (ix) is similar. It follows that Γ_α consists exactly of those sentences of $\mathcal{L}[D(\alpha)]$ which are forced at α, and this is enough (together with the construction of Γ) to establish Veldman's completeness theorem. Informal Markov's Principle is not needed in the proof.

8 ACKNOWLEDGEMENTS

I am grateful to the students of the Graduate Program in Logic and Algorithms (ΜΠΛΑ) in Athens, especially to Garyfallia Vafeiadou and Nikos Vaporis, for reviving and expanding my enthusiasm for this subject over the past few years. Anne Troelstra and Jaap van Oosten unwittingly showed me by example how an encyclopedia article should be written. Mark van Atten, Garyfallia Vafeiadou, Wim Veldman, Richard Vesley, Richard Zach and Yiannis Moschovakis generously

read and commented on versions of the manuscript. I also thank Marcel Guillaume and Paolo Mancosu for providing helpful information. All remaining errors and omissions are exclusively mine.

BIBLIOGRAPHY

[Aczel, 1968] P. Aczel. Saturated intuitionistic theories, in H. Schmidt, K. Schütte and H.-J. Thiele, eds., *Contributions to Mathematical Logic*, North-Holland, Amsterdam, 1968.

[Avigad and Feferman, 1998] J. Avigad and S. Feferman. Gödel's functional ("Dialectica") interpretation, Ch. V of [Buss, 1998], 337-405, 1998.

[Beth, 1956] E. W. Beth. Semantic construction of intuitionistic logic, *Kon. Ned. Akad. v. Wet. Afd. Let. Med. Nieuwe Serie* 19/11, 357-388, 1956.

[Beth, 1959] E. W. Beth. *The Foundations of Mathematics*, North-Holland, Amsterdam, 1959.

[Barzin and Errera, 1927] M. Barzin and A. Errera. Sur la logique de M. Brouwer, *Bull. Acad. Royale Belgique* **13**, 56-71, 1927.

[Beeson, 1985] M. Beeson. *Foundations of Constructive Mathematics*, Springer Verlag, Berlin, 1985.

[Bernays, 1926] P. Bernays. Axiomatische Untersuchung des Aussagenkalküls der 'Principia Mathematica', *Math. Zeitschrift* 25, 305, 1926.

[Bishop, 1967] E. Bishop. *Foundations of Constructive Analysis*, McGraw-Hill, New York, 1967.

[Bridges and Richman, 1987] D. Bridges and F. Richman. *Varieties of Constructive Mathematics*, Cambridge Univ. Press, Cambridge, 1987.

[Brouwer, 1907] L. E. J. Brouwer. *Over de Grondslagen der Wissenschafte*, Thesis, Amsterdam, 1907. English trans. *On the foundations of mathematics* in Heyting, ed., Brouwer [1975, 11-101].

[Brouwer, 1908] L. E. J. Brouwer. *De onbetrouwbaarheid der logische principes*, Tijdschrift voor Wijsbegeerte **2**, 152-158, 1908. English trans. *The unreliability of the logical principles* in Heyting, ed., Brouwer [1975, 107-111].

[Brouwer, 1909] L. E. J. Brouwer. *Het wezen der meetkunde*, Amsterdam, 1909. English trans. *The nature of Geometry* in Heyting, ed., Brouwer [1975, 112-120].

[Brouwer, 1912] L. E. J. Brouwer. *Intuitionism and formalism*, 1912. English trans. by A. Dresden, *Bull. Amer. Math. Soc. 20*, 81-96. Reprinted in Benacerraf and Putnam [1983, 77-89].

[Brouwer, 1918] L. E. J. Brouwer. *Begrundung der Mengenlehre unabhangig vom logischen Satz vom ausgeschlossenen Dritten. Erster Teil: Allgemeine Mengenlehre*, Ver. Kon. Akad. v. Wet. **I, 12**, no. 5, 1918. Reprinted in Heyting, ed., Brouwer [1975, 150-190].

[Brouwer, 1919] L. E. J. Brouwer. *Begrundung der Mengenlehre unabhangig vom logischen Satz vom ausgeschlossenen Dritten. Zweiter Teil: Theorie der Punktmengen*, Ver. Kon. Akad. v. Wet. **I, 12**, no. 7, 1919. Reprinted in Heyting, ed., Brouwer [1975, 191-221].

[Brouwer, 1923] L. E. J. Brouwer. *Intuitionistische splitsing van mathematische grondbegrippen*, Ver. Kon. Akad. v. Wet. **32**, 877-880, 1923. English trans. by W. P. van Stigt in Mancosu [1998, 286-289].

[Brouwer, 1923a] L. E. J. Brouwer. Über die Bedeutung des Satzes vom ausgeschlossen Dritten in der Mathematik, insbesondere in der Funktionentheorie, *Journal für die reine und angewandte Mathematik* **154**, 1-7, 1923. English trans. by S. Bauer-Mengelberg in van Heijenoort [1967, 334-341]. Cf. Brouwer [1954].

[Brouwer, 1923b] L. E. J. Brouwer. *Begründung der Funktionlehre unabhängig vom Logischen Satz der ausgeschlossen Dritten*, Ver. Akad. Wet. Amsterdam **13** No. 2, 1923.

[Brouwer, 1924] L. E. J. Brouwer. Beweis, dass jede volle Funktion gleichmässig stetig ist, Kon. Ned. Akad. v. Wet. Proc. **27**, 189-193, 1924. Reprinted in Heyting, ed., Brouwer [1975, 286-290].

[Brouwer, 1924a] L. E. J. Brouwer. Intuitionistische Zerlegung mathematischer Grundbegriffe, Jber. Deutsch Math. Verein. **33**, 251-256, 1924. Reprinted in Heyting, ed., Brouwer [1975, 275-280]. English trans. by W. P. van Stigt in Mancosu [1998, 290-292].

[Brouwer, 1925] L. E. J. Brouwer. *Zur Begründung der intuitionistische Mathematik I*, Math. Annalen **93**, 244-257, 1925.

[Brouwer, 1927] L. E. J. Brouwer. Über Definitionsbereiche von Funktionen, Math. Annalen **97**, 60-75, 1927. English trans. of §1-3: *On the domains of definition of functions* by S. Bauer-Mengelberg in van Heijenoort [1967, 446-463].

[Brouwer, 1927a] L. E. J. Brouwer. Zur intuitionistische Zerlegung mathematischer Grundbegriffe, Jber. Deutsch Math. Verein. **36**, 127-129, 1927. Reprinted in Heyting, ed., Brouwer [1975, 295-297]. English trans. by W. P. van Stigt in Mancosu [1998, 293-295].

[Brouwer, 1928] L. E. J. Brouwer. Intuitionistische Betrachtungen über den Formalismus, KNAW Proc. **31**, 374-379, 1928. Also Sitzber. Preuss. Akad. Wiss. (1928), 48-52. English trans. of §I by S. Bauer-Mengelberg in van Heijenoort [1967, 490-492].

[Brouwer, 1930] L. E. J. Brouwer. Die Struktur des Kontinuums, 1930. In Heyting, ed., Brouwer [1975, 429-440].

[Brouwer, 1952] L. E. J. Brouwer. Historical background, principles and methods of intuitionism, South-African Jour. Sci. **49**, 139-146, 1952.

[Brouwer, 1954] L. E. J. Brouwer. Addenda and corrigenda and Further addenda and corrigenda to Brouwer [1923a], 1954. English trans. by S. Bauer-Mengelberg, C. Berndes Franck, D. van Dalen and the editor in van Heijenoort [1967, 341-345].

[Brouwer, 1954a] L. E. J. Brouwer. Points and spaces, Canad. Jour. Math. **6**, 1-17, 1954. Reprinted in Heyting, ed., Brouwer [1975, 522-538].

[Brouwer, 1975] L. E. J. Brouwer. *Collected Works, 1*, edited by A. Heyting, North-Holland, Amsterdam, 1975.

[Brouwer, 1981] L. E. J. Brouwer. *Brouwer's Cambridge Lectures on Intuitionism*, edited by D. van Dalen, Cambridge Univ. Press, Cambridge, 1981.

[Buss, 1998] S. Buss, ed. *The Handbook of Proof Theory*, Elsevier, Amsterdam and New York, 1998.

[Church, 1928] A. Church. On the law of excluded middle, Bull. Amer. Math. Soc. **34**, 75-78, 1928.

[Church, 1936] A. Church. An unsolvable problem of elementary number theory, Amer. Jour. Math. **58**, 345-363, 1936.

[Church, 1936a] A. Church. A note on the Entscheidungsproblem, Jour. Symb. Logic **1**, 40-41. *Correction*, Ibid., 101-102, 1936.

[Dedekind, 1888] R. Dedekind. *Was sind und was sollen die Zahlen*, Braunschweig, 1888. English trans. by W. Beman in R. Dedekind [1901], **Essays on the Theory of Numbers**, Open Court, Chicago, reprinted by Dover (1963), 31-105.

[Dragalin, 1988] A. G. Dragalin. Mathematical Intuitionism: Introduction to Proof Theory, Trans. Math. Mon. **67**, Amer. Math. Soc., Providence, R.I., 1988. English trans. by E. Mendelsohn of the 1979 Russian version.

[Dummett, 1977] M. Dummett. *f Elements of Intuitionism*, Clarendon Press, Oxford, 1977.

[Dyson and Kreisel, 1961] V. Dyson and G. Kreisel. *Analysis of Beth's semantic construction of intuitionistic logic*, Stanford Report, 1961.

[Dzik, 2004] W. Dzik. Chains of structurally complete predicate logics with the application of Prucnal's substitution, Reports on Math. Logic **38**, 37-48, 2004.

[Friedman, 1975] H. Friedman. The disjunction property implies the existence property, Proc. Nat. Acad. Sci. **72**, 2877-2878, 1975.

[Gabbay and de Jongh, 1974] D. Gabbay and D. H. J. de Jongh. A sequence of finitely axiomatizable intermediate logics with the disjunction property, Jour. Symb. Logic **39**, 67-78, 1974.

[Gentzen, 1933] G. Gentzen. Über das Verhältnis zwischen intuitionistischer und klassischer Logik, accepted by Math. Annalen, 1933, but withdrawn. Published in Arch. Math. Logik Grund. **16** (1974), 119-132. English trans. in Szabo, ed., Gentzen [1967, 53-67].

[Gentzen, 1935] G. Gentzen. Untersuchungen über das logische Schliessen, Math. Zeitschrift **39** (1934-5), 176-210, 405-431, 1935.

[Gentzen, 1967] G. Gentzen. *Gentzen: Collected Papers*, edited by M. E. Szabo, North-Holland, Amsterdam, 1967.

[Ghilardi, 1999] S. Ghilardi. Unification in intuitionistic logic, Jour. Symb. Logic **64**, 859-880, 1999.

[Glivenko, 1928] V. Glivenko. Sur la logique de M. Brouwer, Bull. Acad. Royale de Belgique **14**, 225-228, 1928.

[Glivenko, 1929] V. Glivenko. Sur quelques points de la logique de M. Brouwer, Bull. Acad. Royale de Belgique **15**, 183-188, 1929. English trans. by A. Rocha in Mancosu [1998, 301-305].

[Gödel, 1932] K. Gödel. Über Vollstandigkeit und Widerspruchfreiheit, Ergebnisse eines math. Koll. **3** (for 1930-1931, pub. 1932), 12-13, 1932.

[Gödel, 1932a] K. Gödel. Zum intuistischen Aussagenkalkül, Akad. Wiss. Wien, Math.-naturwiss. Klasse, Anzeiger **69**, 65-66, 1932.

[Gödel, 1933] K. Gödel. Über Unabhängigkeitsbeweise im Aussagenkalkül, Ergebnisse eines math. Koll. **4** (for 1932, pub. 1933), 9-10, 1933.

[Gödel, 1933] K. Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie, Ergebnisse eines math. Koll., **4** (for 1932, pub. 1933), 34-38, 1933.

[Gödel, 1958] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, Dialectica **12**, 280-287, 1958.

[Harrop, 1956] R. Harrop. On disjunctions and existential statements in intuitionistic systems of logic, Math. Annalen **132**, 347-361, 1956.

[Hessling, 2003] D. Hesseling. *Gnomes in the Fog: The Reception of Brouwer's Intuitionism in the 1920s*, Birkhaüser, Basel, 2003.

[Heyting, 1930] A. Heyting. Sur la logique intuitionniste, Bull. Acad. Royale Belgique **16**, 957-963, 1930. English trans. by A. Rocha in Mancosu [1998, 306-310].

[Heyting, 1930a] A. Heyting. Die formalen Regeln der intuitionistischen Logik, Sitzber. Preuss. Akad. Wiss. (phys.-math. Klasse), Berlin, 42-56, 1930. English trans. in Mancosu [1998, 311-327].

[Heyting, 1930b] A. Heyting. Die formalen Regeln der intuitionistischen Mathematik II, Sitzber. Preuss. Akad. Wiss. (phys.-math. Klasse), Berlin, 57-71, 1930.

[Heyting, 1930c] A. Heyting. Die formalen Regeln der intuitionistischen Mathematik III, Sitzber. Preuss. Akad. Wiss. (phys.-math. Klasse), Berlin, 158-169, 1930.

[Heyting, 1956] A. Heyting. *Intuitionism: An Introduction*, North-Holland, Amsterdam, 1956.

[Heyting, 1946] A. Heyting. On weakened quantification, *Jour. Symb. Logic*, **11**, 119-121, 1946.

[Heyting, 1961] A. Heyting. Infinitistic methods from a finitist point of view, in *Infinitistic Methods*, Proceedings of the Symposium on Foundations of Mathematics (Warsaw 1959), Pergamon Press, 185-192, 1961.

[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*, Springer Verlag, Berlin, 1928.

[Hilbert and Bernays, 1934] D. Hilbert and P. Bernays. *Grundlagen der Mathematik 1*, Springer Verlag, Berlin, 1934.

[Hilbert and Bernays, 1939] D. Hilbert and P. Bernays. *Grundlagen der Mathematik 2*, Springer Verlag, Berlin, 1939.

[Iemhoff, 2001] R. Iemhoff. On the admissible rules of intuitionistic propositional logic, Jour. Symb. Logic **66**, 281-294, 2001.

[Iemhoff, 2005] R. Iemhoff. Intermediate logics and Visser's rules, Notre Dame Jour. Form. Logic **46**, 65-81, 2005.

[de Jongh, 1970] D. H. J. de Jongh. The maximality of the intuitionistic propositional calculus with respect to Heying's arithmetic, Jour. Symb. Logic **36**, 606, 1970.

[Kleene, 1945] S. C. Kleene. On the interpretation of intuitionistic number theory, Jour. Symb. Logic **10**, 109-124, 1945.

[Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics*, North-Holland, Amsterdam, 1952.

[Kleene, 1962] S. C. Kleene. Disjunction and implication under implication in elementary intuitionistic formalisms, Jour. Symb. Logic **27**, 11-18, 1962. Addendum (1963) in Jour. Symb. Logic **28**, 154-156.

[Kleene, 1965] S. C. Kleene. Classical extensions of intuitionistic mathematics, in Y. Bar-Hillel, ed., *Logic, Methodology and Philosophy of Science*, North-Holland, Amsterdam, 31-44, 1965.

[Kleene, 1967] S. C. Kleene. Constructive functions in "The Foundations of Intuitionistic Mathematics", in *Logic, Methodology and Philos. of Science III*, North-Holland, Amsterdam, 137-144, 1967.

[Kleene, 1969] S. C. Kleene. *Formalized recursive functionals and formalized realizability*, Mem. Amer. Math. Soc. **89**, 1969.

[Kleene and Vesley, 1965] S. C. Kleene and R. E. Vesley. *The Foundations of Intuitionistic Mathematics, Especially in Relation to Recursive Functions*, North-Holland, Amsterdam, 1965.

[Kolmogorov, 1925] A. Kolmogorov. O principe tertium non datur, Matematiceskij Sbornik **32**, 646-667, 1925. English trans. in van Heijenoort [1967, 414-437].

[Kolmogorov, 1932] A. Kolmogorov. Zur Deutung der intuitionistische Logik, Math. Zeitschrift **35**, 58-65, 1932. English trans. in Mancosu [1998, 328-334].

[Krauss, 1992] P. Krauss. A constructive interpretation of classical mathematics, Math. Schriften Kassel, Preprint No. 5/92, 1992.

[Kreisel, 1958] G. Kreisel. Mathematical significance of consistency proofs, Jour. Symb. Logic **23**, 155-181, 1958.

[Kreisel, 1959] G. Kreisel. The non-derivability of $\neg\forall x A(x) \rightarrow \exists x \neg A(x)$, A primitive recursive, in intuitionistic formal systems (abstract), Jour. Symb. Logic **23** (No. 4 for 1958, publ. 1959), 456-457.

[Kreisel, 1962] G. Kreisel. On weak completeness of intuitionistic predicate logic, Jour. Symb. Logic **27**, 139-158, 1962.

[Kreisel, 1968] G. Kreisel. Lawless sequences of natural numbers, Comp. Math. **20**, 222-248, 1968.

[Kreisel and Troelstra, 1970] G. Kreisel and A. S. Troelstra. Formal systems for some branches of intuitionistic analysis, Ann. Math. Logic **1**, 229-387, 1970.

[Kripke, 1965] S. Kripke. Semantical analysis of intuitionistic logic, in *Formal Systems and Recursive Functions* (J. Crossley and M. Dummett, eds.), North-Holland, Amsterdam, 92-130, 1965.

[Krol', 1973] M. D. Krol'. A topological model for intuitionistic analysis with Kripke's Schema, Zeitsch. math. Logik Grund. **24**, 427-436, 1973.

[Loeb, 2005] I. Loeb. Equivalents of the (Weak) Fan Theorem, *Ann. Pure and Appl. Logic*, **132**, 51-66, 2005.

[Mancosu, 1998] P. Mancosu. *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s*, Oxford Univ. Press, Oxford and New York, 1998.

[Mancosu and van Stigt, 1998] P. Mancosu and W. van Stigt. Intuitionistic logic, in Mancosu [1998, 275-285].

[Mancosu *et al.*, 2009] P. Mancosu, R. Zach and C. Badesa. The development of mathematical logic from Russell to Tarski, in L. Haaparanta, ed., *The Development of Modern Logic: A Philosophical and Historical Handbook*, Oxford University Press, Oxford, 2009.

[Martin-Löf, 1984] P. Martin-Löf. *Intuitionistic Type Theory*, Bibliopolis, Naples, 1984.

[Mints, 1976] G. Mints. Derivability of admissible rules, *Jour. Soviet Math.*, **6**, 417-421, 1976.

[Moschovakis, 1971] J. R. Moschovakis. Can there be no non-recursive functions?, Jour. Symb. Logic **36**, 309-315, 1971.

[Moschovakis, 1996] J. R. Moschovakis. A classical view of the intuitionistic continuum, Ann. Pure and Appl. Logic **81**, 9-24, 1996.

[Moschovakis, 2003] J. R. Moschovakis. Classical and constructive hierarchies in extended intuitionistic mathematics, Jour. Symb. Logic **68**, 1015-1043, 2003.

[Moschovakis, 2004] J. R. Moschovakis. Intuitionistic Logic, in the Stanford On-Line Encyclopedia of Philosophy, 2004. {`http://www.plato.stanford.edu`}.

[Mostowski, 1948] A. Mostowski. Proofs of non-deducibility in intuitionistic functional calculus, Jour. Symb. Logic **13**, 204-207, 1958.

[Myhill *et al.*, 1970] J. Myhill, A. Kino and R. Vesley, eds. *Intuitionism and Proof Theory*, North-Holland, Amsterdam, 1970.

[Nelson, 1947] D. Nelson. Recursive functions and intuitionistic number theory, Trans. Amer. Math. Soc. **61**, 307-368, 1947.

[Presburger, 1930] M. Presburger. Über die Vollstandigkeit eines gewissen Systems der Arithmetik ganzer Zahlen in Welchem die Addition als einzige Operation hervortritt, Sprawozdanie z I Kongresu Matematyków Krajów Slowiańskich (Comptes-rendus du I Congrés des Mathématiciens des Pays Slaves), Warszawa 1929, Warsaw 92-101, 395, 1930.

[Rasiowa, 1955] H. Rasiowa. Algebraic models of axiomatic theories, Fund. Math. **41**, 291-310, 1955.

[Rasiowa and Sikorski, 1954] H. Rasiowa and R. Sikorski. On existential theorems in non-classical functional calculi, Fund. Math. **41**, 21-28, 1954.

[Rasiowa and Sikorski, 1963] H. Rasiowa and R. Sikorski. *The Mathematics of Metamathematics*, Warsaw, 1963.

[Rosser, 1936] J. B. Rosser. Extensions of some theorems of Gödel and Church, Jour. Symb. Logic **1**, 87-91, 1936.

[Rybakov, 1997] V. Rybakov. *Admissibility of Logical Inference Rules*, Elsevier, Amsterdam, 1997.

[Schütte, 1950] H. Schütte. Schlussweisen-Kalküle der Prädikatenlogik, Math. Annalen **122**, 47-65, 1950.

[Scott, 1968] D. Scott. Extending the topological interpretation to intuitionistic analysis, Comp. Math. **20**, 194-210, 1968.

[Scott, 1970] D. Scott. Extending the topological interpretation to intuitionistic analysis II, in [Myhill *et al.*, 1970], 235-255, 1970.

[Scott, 1979] D. Scott. Identity and existence in intuitionistic logic, in *Applications of Sheaves* (M. Fourman, C. Mulvey and D. Scott, eds.), Springer Verlag, Berlin, 660-696, 1979.

[Smorynski, 1973] C. Smorynski. Applications of Kripke models, Ch. V of Troelstra, ed. [Troelstra, 1973], 324-391, 1973.

[Spector, 1962] C. Spector. Provably recursive functionals of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics, in J. C. E. Dekker, ed., *Recursive Function Theory*, AMS Proc. Symp. Pure Math V., Providence, 1-27, 1962.

[Troelstra, 1968] A. S. Troelstra. The scientific work of A. Heyting, Comp. Math. **20**, 3-12, 1968.

[Troelstra, 1971] A. S. Troelstra. Notions of realizability for intuitionistic arithmetic and intuitionistic arithmetic in all finite types, in J. Fenstad, ed. *The Second Scandinavian Logic Symposium*, North-Holland, Amsterdam, 369-405, 1971.

[Troelstra, 1973] A. S. Troelstra, ed. *Metamathematical Investigations of Intuitionistic Arithmetic and Analysis*, Springer Lecture Notes in Math. **344**, Springer Verlag, Berlin, 1973.

[Troelstra, 1973a] A. S. Troelstra. Notes on intuitionistic second order arithmetic, in A. Mathias and H. Rogers, eds. [1973], *Cambridge Summer School in Mathematical Logic*, Springer, Berlin, 171-205, 1973.

[Troelstra, 1974] A. S. Troelstra. Note on the fan theorem, Jour. Symb. Logic **39**, 584-596, 1974.

[Troelstra, 1977] A. S. Troelstra. *Choice Sequences, a Chapter of Intuitionistic Mathematics*, Clarendon Press, Oxford, 1977.

[Troelstra, 1991] A. S. Troelstra. History of constructivism in the twentieth century, ITLI (Inst. v. Taal, Logica en Informatie) Prepublication Series ML-91-05, Amsterdam, 1991.

[Troelstra, 1998] A. S. Troelstra. Realizability, Ch. VI of [Buss, 1998], 407-473, 1998.

[Troelstra and van Dalen, 1988] A. S. Troelstra and D. van Dalen. *Constructivism in Mathematics: An Introduction*, **1** and **2**, North-Holland, Amsterdam, 1988.

[van Atten, 2005] M. van Atten. Luitzen Egbertus Jan Brouwer, in the Stanford On-Line Encyclopedia of Philosophy, 2005. `http://www.plato.stanford.edu`

[van Atten and Kennedy, 2007] M. van Atten and J. Kennedy. Gödel's Logic, in this volume.

[van Dalen, 1984] D. van Dalen. How to glue analysis models, Jour. Symb. Logic **49**, 1339-1349, 1984.

[van Dalen, 1990] D. van Dalen. The war of the frogs and the mice, or the crisis of the Mathematische Annalen, Math. Intelligencer **12(4)**, 17-31, 1990.

[van Dantzig, 1947] D. van Dantzig. On the principles of intuitionistic and affirmative mathematics, Ver. Kon. Akad. v. Wet. **50**, 918-929, 1092-1103; also Indagationes math. **9**, 429-440, 506-517, 1947.

[van Heijenoort, 1967] J. van Heijenoort. *From Frege to Gödel: A Source Book in Mathematical Logic 1897-1931*, Harvard Univ. Press, Cambridge, Mass, 1967.

[van Oosten, 1991] J. van Oosten. A semantical proof of de Jongh's theorem, Arch. Math. Logic **31**, 105-114, 1991.

[van Oosten, 2002] J. van Oosten. Realizability: a historical essay, Math. Struct. Comp. Sci. **12**, 239-263, 2002.

[van Stigt, 1990] W. P. van Stigt. *Brouwer's Intuitionism*, North-Holland, Amsterdam, 1990.

[Veldman, 1976] W. Veldman. An intuitionistic completeness theorem for intuitionistic predicate logic, Jour. Symb. Logic **41(1)**, 159-166, 1976.

[Vesley, 1970] R. E. Vesley. A palatable substitute for to Kripke's Schema, in [Myhill *et al.*, 1970].

[Vesley, 1980] R. E. Vesley. Intuitionistic analysis: The search for axiomatization and understanding, in *The Kleene Symposium*, North- Holland, Amsterdam, 317-331, 1980.
[Visser, 1999] A. Visser. Rules and arithmetics, Notre Dame Jour. Form. Logic **40**, 116-140, 1999.
[von Plato, 2007] J. von Plato. Gentzen's Logic, in this volume.
[Wavre, 1926] R. Wavre. Logique formelle et logique empiriste, Revue de Métaphysique et de Morale **33**, 65-75, 1926.
[Zach, 1999] R. Zach. Completeness before Post, Bull. Symb. Logic **5(3)**, 303-330, 1999.

THORALF ALBERT SKOLEM

Jens Erik Fenstad and Hao Wang

1 INTRODUCTION AND ACKNOWLEDGEMENT

This chapter is based on the corresponding sections of Skolem's *Selected Works in Logic* [Skolem, 1970]. The biographical part is a revised version of the chapter *Thoralf Albert Skolem in memoriam* [Skolem, 1970, pp. 9–15]. The section on Skolem's work in logic is a reprint of the corresponding chapter in the Selected Works (Hao Wang, *A survey of Skolem's work in logic*, [Skolem, 1970, pp. 17–52]). Wang's survey, written almost 40 years ago, is still an excellent introduction to Skolem's work. But it needs to be updated; it also needs some factual corrections. I have, therefore, added some supplementary notes and references to the current discussion of Skolem's work. The literature on Skolem is vast; I have tried to draw attention to some of the more important sources. I have also, in particular in Section 4.6, added a few "interpretative remarks" expressing a personal view on Skolem and his discussion of foundational matters.

The bibliography in Skolem 1970 is a complete listing of all the scientific works of Skolem; here we have included only those items that relate to his work in logic. The bibliography in Skolem 1970 was prepared by Herman Ruge Jervell. Jervell has also been of great help in the preparation of this chapter. I want to thank him for expert technical assistance and valuable advice on the scientific content.

2 BIOGRAPHY OF SKOLEM

Professor Thoralf Albert Skolem died on 23 March 1963, almost 76 years old. His death was unexpected. He seemed to be as active as ever. He was preparing for a visit to the USA where, as on many previous occasions, he had been invited to lecture at several universities. Age had not diminished either his research drive or creative ability.

Thoralf Skolem was born on 23 May 1887 in Sandsvær, in the county of Buskerud in Southern Norway. His parents were Helene Olette Vaal (1850–1914) and Even Skolem (1849–1926), who taught in elementary school, but the family was mainly farming people. Skolem passed his "Examen artium" (which is the concluding examination of the Norwegian gymnasium) in 1905 and immediately started to study mathematics and science at the University of Oslo. The study for the state examination (embetseksamen) was at the time divided into two parts, a first or

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

general part covering a large number of subjects, and a higher or specialized part in one subject, including a thesis.

Skolem passed the general part of the state examination in 1909 and the full state examination in 1913, specializing in mathematics. For his thesis, *Undersøkelser innenfor logikkens algebra (Investigations on the Algebra of Logic)* he obtained the best possible mark. His outstanding achievement was reported, as happens on rare occasions, to the King of Norway.

Already as a student, after he had passed the general part of the state examination, Skolem became in 1909 the private assistant to the well-known physicist Kristian Birkeland. He served in this position to 1914, and it was in physics that Skolem, in collaboration with Birkeland, published his first scientific papers. Skolem accompanied Birkeland on an expedition to Egypt and Sudan in 1913–1914 to observe the zodiacal light.

In the winter of 1915–1916 Skolem studied in Göttingen, experiencing fully the difficult living conditions caused by the First World War. From 1916 to 1918 he was a Research Fellow at the University of Oslo, and in 1918 he was appointed to a newly created position as docent of Mathematics at the same university.

The docent position at that time was a permanent position and functioned in a certain sense as a professorship-in-waiting. It carried a significantly lower salary than that of a professorship, but required the holder to give only three lectures per week and spend the rest of his time on his own research — no administration foreseen. In current terms it was a permanent postdoc. It was expected that the docent would become the next professor when a vacancy occurred. And despite the permanency of the position, a docent who did not advance to a professorship was definitely an academic failure. When Skolem was appointed docent in 1918 at the age of 31, he had only published two papers in his own name, but it was evident that he was judged to be a coming man in Norwegian mathematics.

In 1926, when he was nearly 40 years old, Skolem obtained his doctorate, the thesis entitled *Einige Sätze über ganzzahlige Lösungen gewisser Gleichungen und Ungleichungen.* The somewhat advanced age has the following explanation. In their younger years, Viggo Brun and Skolem agreed that neither of them would bother to obtain the degree of Doctor, probably feeling that, in Norway, it served no useful function in the education of a young scientist. But in the middle twenties a younger generation of Norwegian mathematicians emerged. It seems that Skolem then felt he too ought to fulfill the formal requirement of having a doctorate, and he "obtained permission" from Brun to submit a thesis. In 1924 Brun had become a professor in mathematics at the Norwegian Institute of Technology.

On 23 May 1927 Skolem married Edith Wilhelmine Hasvold.

From 1930 to 1938 Skolem was a Research Associate at the Chr. Michelsen's Institute in Bergen. This was a very independent position, equivalent in many respects to a professorship, but not carrying any teaching or administrative duties. One of the conditions of the research grant was that he should live in Bergen. Since there was no university in Bergen at that time, this meant that he was scientifically very isolated. He also had difficulties in obtaining the necessary

mathematical literature while living there, which can be inferred from the preface to the monograph *Diophantische Gleichungen*, which he wrote during his Bergen years.

In 1938, at the age of 51, Skolem was appointed professor at the University of Oslo. He conducted the regular graduate courses in algebra and number theory, and rather infrequently lectured on mathematical logic. On a few occasions he acted as thesis advisor for the state examination, the case of Ø. Ore being the most prominent example. Thoralf Skolem was very modest and retiring by nature. He did not create any school and had no research students, but through his great accomplishments and research drive he inspired more than one of the younger Norwegian mathematicians.

Skolem retired in 1957, thereafter visiting American universities on several occasions. Professor Skolem always participated actively in the various administrative and organizational duties which is so necessary in any field of science. He was for many years President of the Norwegian Mathematical Association. For a long period he was Editor of *Norsk matematisk tidsskrift*, and continued after the reorganization as Editor of *Mathematica Scandinavica* from 1952. He also served as Associate Editor for many journals. Professor Skolem was a member of several scientific academies, and as early as 1918 became a member of Vitenskaps-Akademiet in Oslo. He was frequently invited to lecture at international congresses and symposia, and received many scientific honours, e.g. in 1962 he was awarded the Gunnerus Medal by Det Kongelige Norske Vitenskabers Selskab. In 1954 he was named a Knight of the 1st Class in the Royal Order of St. Olav by the King of Norway.

It is for his scientific accomplishments, however, that Professor Skolem will be remembered. Teaching and organizational activities were necessary duties, research was his life.

A rough count reveals that he wrote close to 200 papers, some of them opening up new fields of research, most of them of lasting scientific value, and a few of perhaps lesser interest today; for instance, as Editor of *Norsk matematisk tidsskrift*, he would write a paper or two to secure a complete issue as the publication date approached. Some notes on algebra and number theory belong in this category. In the same journal there is the somewhat curious paper *En liten studie i transfinit mekanikk* (*A small study on transfinite mechanics*) from 1940. But he also wrote some very important papers in this rather obscure journal, e.g. *Über die mathematische Logik* in 1928. Professor Skolem also wrote many reviews here, not just within the field of mathematics, but also in physics and philosophy, subjects which strongly held his interest. Logician and number-theorist, he was one of the very few Norwegian mathematicians to regularly attend the physicists' meetings.

Professor Skolem published the greater part of his papers in *Videnskapsakademiets Skrifter* and *Avhandlinger*, in *Norsk matematisk tidsskrift*, and in *Det Kongelige Norske Videnskabsselskabs Forhandlinger*. Another rough count shows that approximately two-thirds of his papers are published here, which has often been of great inconvenience to his fellow scientists abroad. The proceedings of an academy,

being a haphazard mixture of every field of science, are very often stacked away in some central library, and rarely seen in the special library of a mathematical institute. This means that many of Professor Skolem's most important contributions have been difficult to obtain.

Great mathematicians start to publish young. Professor Skolem was a counterexample. When appointed docent in 1918, 31 years old, he had written only two papers (disregarding the notes he co-authored with Kr. Birkeland), but from 1920 on he published a series of papers which gave him instant recognition as one of the foremost among logicians: *Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen* dates from 1919. The year after, *Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit und Beweisbarkeit mathematischen Sätze nebst einem Theoreme über dichte Mengen* appeared. In 1922 he gave a lecture at the 5th Scandinavian Mathematics Congress *Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre*, and finally in 1923 the important paper *Begründung der elementären Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnugsbereich* was published. These are classics in the field of mathematical logic, although they remained to a large extent unnoticed and unread by contemporary mathematicians, in particular Scandinavian mathematicians!

It would be impossible to characterize the scientific work of Professor Skolem in just a few words, but we feel that it is essential to articulate a few of the ideas and some of the problems he worked on.

In the paper from 1920 Skolem proved the theorem which is now known as the *Löwenheim–Skolem* theorem. Briefly, this states that if a finite or countable infinite set of sentences formalized within a first-order predicate calculus is satisfiable (or, in other terminology, has a model), then the sentences are satisfiable within a countable domain. A by-product of the proof is Skolem's well-known normal form for the predicate calculus. In his early papers we also find the technique of using *Skolem functions* to eliminate quantifiers and in the construction of models — a technique of extraordinary importance in contemporary general model theory.

In the 1922 lecture the Löwenheim–Skolem theorem was applied to a formalization of set theory. The result was a relativization of the notion of set, later known as the Skolem paradox: If the axiomatic system (e.g. as presented by Zermelo with Skolem's notion of definiteness) is consistent, i.e. if it is at all satisfiable, then it must be satisfiable within a countable *Denkbereich* (domain). But does this not contradict Cantor's theorem of the uncountable, the existence of a never-ending sequence of transfinite powers? We shall discuss the "paradox" at length in later sections of this paper, and what it means for for the foundation of set theory. Here we just point out, as Skolem did in his 1922 lecture, that concepts introduced in an axiomatic system only seem to acquire a meaning relative to a model or interpretation of the system. Thus let A be a set defined and proved to be uncountable by the axioms, and let M by the Löwenheim–Skolem theorem be a countable model of the axioms. Let further A_M be the interpretation of A

in M. Then A_M is uncountable in the sense of the model M, i.e. there is no function inside M that establishes a one-to-one correspondence between A_M and the natural numbers in M. But A_M as a part of the countable model M, must be countable.

In the same lecture Skolem made several other important contributions to the foundation of set theory. He came up with a method, now universally accepted, for eliminating the psychologism inherent in Cantor's description of a set and still present in the axiomatization given by Zermelo. Skolem also extended and thus completed Zermelo's system on an important point by introducing the axiom of replacement. This was also done at about the same time by A. Fraenkel. The Fraenkel contributions became better known, but in all fairness one might claim that what is today called the *Zermelo–Fraenkel* system of axiomatic theory ought really to be known as the *Zermelo–Fraenkel–Skolem* system.

Skolem himself did not have much confidence in set theory as a foundation for "real" mathematics, and he was extremely doubtful about the transfinite powers and non-constructive modes of reasoning of set-theoretical mathematics. His own preferences are better represented by the important 1923 paper. Here Skolem tries to build up elementary arithmetic without applying the unrestricted quantifiers "for all" and "there exists" to the infinite completed totality of natural numbers. Historically, this is perhaps "a first" paper in the theory of recursive arithmetic. In a paper from 1924 he also gave a constructive proof for the fundamental theorem of algebra. But, whereas recursion theory has witnessed a rich development, the strictly finitistic construction of mathematics advocated by Skolem has aroused less interest. In a lecture to the 1950 International Congress of Mathematicians he hoped that "the very natural feature of my considerations would convince people that this finitistic treatment of mathematics was not only a possible one but the true or correct one". For an ordinary mathematician this is an extreme position, so Skolem hastens to add: "Now I will not be misunderstood. I am no fanatic." To this, everyone who knew Skolem would surely agree.

Skolem's work in logic did not create much interest among his Scandinavian colleagues at the time, and later he indicated that he could not derive much inspiration while his papers remained unread. So, from the beginning of the 1920s, he turned to more traditional and "respectable" fields — algebra and number theory. Skolem has contributed extensively to the theory of Diophantine equations, not so much to the study of particular equations, but in giving results and methods of a more general nature. He is probably best known for the so-called p-adic method, perhaps one of the more significant developments in Diophantine equations in later years. The method can be used quite generally to conclude that certain equations have only a finite number of solutions in rational integers, and in many cases it is possible to give constructive methods for actually finding the solutions. An important paper by Skolem on this method is *Einige Sätze über p-adische Potenzreihen mit Anwendung auf gewisse exponentielle Gleichungen* from 1935. It is not possible here to enter into the details of the method. The reader may refer to an excellent survey by Skolem himself, *The use of p-adic methods in the theory*

of Diophantine equations from 1954. In 1938 Skolem wrote a book *Diophantische Gleichungen* in the Ergebnisse series.

Skolem has also written a large number of papers on algebra. He is perhaps best known in this field for a result published in 1927 in *Zur Theorie der assoziativen Zahlensysteme*. The theorem was later rediscovered by Emmy Noether and R. Brauer, and to-day carries the name *Skolem–Noether*. It characterizes the automorphisms of simple algebras and says in its modern form that there are no other but inner automorphisms of a simple algebra A of finite rank over a field K, where K is the centre of A and supposed invariant under the automorphisms.

Skolem's contribution to algebra and number theory, which is much more extensive than indicated here (e.g. there is a paper from 1944 *Über Nebenkörper und Nebenringe* which has not perhaps been sufficiently exploited), suffices to give him an honourable position in the Mathematicians' hierarchy. In addition, there is his work on logic.

In 1928 he published a lecture *Über die mathematische Logik* which he gave earlier the same year at a meeting of *Norsk Matematisk Forening*. This rather modest title conceals a very important paper. He sets himself the task of showing how the deduction problem for elementary logic can be reduced to an arithmetical-combinatorial problem. Generally, the latter problem is not effectively solvable, but for certain classes of formulas it is possible to obtain an effective decision procedure. The examples given by Skolem are important contributions to the *Entscheidungsproblem* of logic. Skolem wrote several papers on this topic, and, of necessity, they are quite combinatorial in method. Typical in this respect is a short paper from 1933 *Ein kombinatorischer Satz mit Anwendung auf ein logisches Entscheidungsproblem*. The combinatorial theorem referred to is the well-known Ramsay's theorem. Skolem gave a slightly sharper version with a simpler proof.

Skolem has written many papers on pure combinatorics. In collaboration with Viggo Brun he edited the second edition of Netto's *Lehrbuch der Combinatorik*, which was greatly enriched with many original contributions by Brun and Skolem.

Skolem's work in combinatorics is still an active part of current research. But rather than giving a report on this, we will mention the sculpture, *Skolem, choc de blocs & chiffres au vent*, by Jessica Stockholder in the garden of the Institut des Hautes Études Scientifique in Paris. The sculpture has its origin in a project for school children and is based on a combinatorial problem originally proposed by Skolem.

In 1929 one of the better known of Skolem's contributions to logic appeared, *Über einige Grundlagenfragen der Mathematik*. In it he returns to the proof of the Löwenheim-Skolem theorem and gives two new ones. One is short, but uses the axiom of choice. The second, which is not constructive either, deserves special comment. Skolem did not think of logic as an axiomatically given science. For him the notion of consistency was to be interpreted semantically, i.e. as satisfiability in some non-empty domain of individuals. Today, it is more common to formulate the notion proof-theoretically: there is no formula A such that both A and not-A are provable. When Skolem stated the Löwenheim-Skolem theorem in the following

manner: "If a formula A is consistent, then it is satisfiable in a countable domain", one must remember that it is the semantic interpretation of "consistency" which is intended.

On the proof-theoretic interpretation one obtains quite a different theorem: Gödel's completeness theorem for predicate logic. But what merits special comment is the fact that the second 1929 proof of Skolem can, with some additions, be made to serve as proof of Gödel's theorem.

The axiom of choice received an elegant formulation in the 1929 paper. There exists a "global" choice function w such that $x \in y \rightarrow w(x) \in y$. Skolem deeply mistrusted this axiom. In a lecture from 1932 he stated: "If one works within a completely formalized mathematics, based on a finite number of precisely stated axioms, there is nothing to discuss but questions of consistency and ease of manipulation. But in ordinary mathematical practice, e.g., in the usual studies on continua which are never given by a set of specified formal rules, the axiom of choice is, in my opinion, definitely undesirable–a kind of scientific fraud."

Towards the end of the 1929 paper Skolem expressed some doubts about the complete axiomatizability of mathematical concepts. His scepticism was based on the set-theoretic relativism which follows from the Löwenheim-Skolem theorem. In 1929 he could give only some partial results, but in a paper from 1934 (and a previous one from 1933) *Über die Nichtcharacterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen* he could prove that there is no finite or countably infinite set of sentences in the language of Peano arithmetic which characterizes the natural numbers. Today, this follows as a simple consequence of Gödel's completeness theorem. The technique used by Skolem was a more direct model-theoretic construction. And this technique, suitably refined to the so-called *ultraproduct* construction, has been an important tool in recent work on model theory.

Skolem has also written many papers on recursion theory since 1923. Among his many contributions may be mentioned a short and elegant paper from 1940, *Einfacher Beweis der Unmöglichkeit eines allgemeinen Lösungsverfahren für Arithmetische Probleme*. He first shows that every general recursive relation is arithmetical and concludes from this, by a simple application of Cantor's diagonal method, that if one identifies "effective" with general recursive, then there is no effective decision method for arithmetical problems. Perhaps the main unsolved problem in this field was at that time Hilbert's 10th problem, which postulates an effective method for solving Diophantine equations. It is somewhat strange that Skolem, being equally at home in recursion theory and Diophantine equations, never seemed to have worked on this problem. In a paper from 1962 he remarks: "It has been asked whether the recursively enumerable sets are all of them Diophantine sets ... I regret not having had the opportunity to study this question seriously."

As previously remarked, Skolem published most of his papers in Norwegian journals, and they have not always been easy to obtain for mathematicians abroad. Others have thus rediscovered his results. One example is the Skolem-Noether

theorem. Another is Skolem's contributions to lattice theory, which remained completely unnoticed at the time of their publication. In 1936 he even had to write a paper *Über gewisse 'Verbände' oder 'Lattices'* giving a survey of results which he had obtained in 1913 and 1919 and which were rediscovered in the 1930s. Skolem had been the first, for example, to determine the free distributive lattice generated by n elements [1913]; he had shown that every implicative lattice is distributive and, as a partical converse, that every finite distributive lattice is implicative [1919]. These results are of great interest for the algebraization of mathematical logic.

Age did not diminish Skolem's research activities. From 1948 to 1957, i.e. from his 61st to his 70th year, he published 48 papers. Skolem was a truly creative mathematician. In his greatest papers he started out from simple, but, as we now recognize, fundamental situations. What he then wrote has remained basic for all further work.

3 A SURVEY OF SKOLEM'S WORK IN LOGIC

Skolem has made many fundamental contributions to modern investigations on the foundations of mathematics. If one has to single out one most intriguing item, it would probably be his work on non-standard models of set theory and number theory. He has also contributed several other basic new ideas and methods. A comprehensive unifying theme is the nature of quantifiers (with infinite ranges):

1. The Skolem functions and countable models as a general method of analysing quantifiers both in pure logic and in axiomatic theories.

2. The founding of recursive arithmetic as a method of developing mathematics on a quantifier-free basis.

3. The invention and powerful applications of the decision method of eliminating quantifiers for axiomatic theories.

4. Various results on the decision and reduction problems of quantification theory (e.g. the Skolem normal form).

5. The classic explication of Zermelo's concept of "definite property" by the notation of quantification theory.

6. The discovery and emphasis on non-standard models of axiomatic number theory and set theory.

In addition Skolem has made contributions to the algebraisation of logic (e.g. the "Skolem lattices"). In contemporary terminology, Skolem's main contributions are to model theory, recursive function theory, axiomatic set theory, and decision problems. This way of describing is, however, rather misleading since his work dealt with mathematical logic before such fragmentations became prevalent and

has a quite different flavor from the more abstract current approaches: rather he participated in the founding of these subjects by introducing initially several of the basic ideas in their concrete and naked form.

Skolem has a tendency of treating general problems by concrete examples. Often proofs seem to be presented in the same order as he came to discover them. This results in a fresh informality as well as a certain inconclusiveness. Many of his papers give the impression of reports on work in progress. Yet his ideas are often pregnant and potentially capable of wide applications. He was very much a "free spirit": he did not belong to any school, he did not found any school of his own, he did not usually make heavy use of known results in more specialized developments, rather he was very much an innovator and most of his papers can be read and understood by people without much specialized knowledge. It seems quite likely that if he were young today, logic, in its more developed stage, would not have appealed to him.

In Skolem's earlier work, he was much influenced by Schröder's books and Löwenheim's paper of 1915. He was also familiar with Zermelo's work in axiomatic set theory, and stated explicitly that he had studied Whitehead and Russell before the autumn of 1919. His concurrent interest in ordinary number theory was on several occasions helpful, notably in the invention of the necessary combinatorial lemmas. During the latter half of his career, he was not always successful in taking fully into consideration known results related to what he was working on.

In terms of the dates of publication, Skolem's principal work in logic appeared during the 15-year period 1919–1933 (when he was 32 to 46); in particular, the eight papers [1919a; 1920a; 1922b; 1923a; 1928a; 1929b; 1930c; 1933d]. Ideas for some of theses papers can be traced back to 1915-18. There are many other interesting papers, but they do not seem to be equally central to logic and foundations. The publication from 1934 onwards contain mostly remarks, exposistions, alternative proofs of known theorems and partial results towards higher goals.

In what follows, we shall consider Skolem's work in logic under four general headings: quantification theory, axiomatic set theory, algebraic aspects, and recursive functions.

I am grateful to Professors Paul Bernays and Kurt Gödel for reading an earlier draft of this paper and pointing out several errors and inaccuracies.

3.1 Quantification theory (Elementary logic)

3.1.1 Countable models and completeness

This section is concerned with §1 of [1920a], §3 of [1922b], [1928a], §4 and §5 of [1929b], §1 of [1941a], and [1955c]. Related materials includes the first edition of Hilbert–Ackermann, Herbrand's dissertation, Gödel's dissertation (and the published paper on the completeness of elementary logic), and the second volume of Hilbert–Bernays.

A quantificational schema is obtained from atomic schemata (predicate letters followed by variables) by truth-functional connectives and quantifiers in the famil-

iar manner. One can consider the problems of satisfiability and models (in particular, countable models) of any such schema without appeal to a formal system for proving or refuting such schemata. In contrast, the problem of completeness cannot be studied except by reference to a formal system since completeness means that all desired schemata (viz. the valid ones) are *provable in a formal system* (of quantification theory).

In §1 of [1920a], Skolem proves several basic results without reference to any formal system of logic.

THEOREM 1 The Skolem normal form theorem. *There is a general method by which, given any quantificational schema W, we can find another one W' in the Skolem satisfiablity normal form*

(1) $\forall x_1 \ldots \forall x_m \exists y_1 \ldots \exists y_n U(x_1, \ldots, x_m, y_1, \ldots, y_n)$

such that W is satisfiable if and only if W' is.

The proof of this can be illustrated by a simple example:

(2) $\forall x \exists u \forall y U(x, u, y)$

Take a new letter A. It is then clear that (2) is satisfiable in a given domain if and only if the following is:

(3) $\forall x \exists w A(x, w) \wedge \forall x \forall u (A(x, u) \equiv \forall y U(x, u, y))$

Hence, if and only if the following is

$$\forall x \exists w A(x, w) \wedge \forall x \forall u ((A(x, u) \wedge \forall y U(x, u, y)) \vee (\neg A(x, u) \wedge \exists z \neg U(x, u, z))$$

Therefore, (2) is satisfiable in a domain if and only if the following is:

(4) $\forall x \forall u \forall y \exists w \exists z (A(x, w) \wedge ((A(x, u) \wedge U(x, u, y)) \vee (\neg A(x, u) \wedge U(x, u, z))))$

This illustrates Skolem's rather intuitive proof of Theorem 1. It differs from the proof in the standard textbooks which goes back to Gödel's dissertation either directly or through the second edition of Hilbert–Ackermann. The alternative is more economical on quantifiers. Thus, we replace $\equiv$ by $\supset$ in (3) and arrive at:

(5) $\forall x \exists w A(x, w) \wedge \forall x \forall u (A(x, u) \supset \forall y U(x, u, y))$

(6) $\forall x \exists w A(x, w) \wedge \forall x \forall u \forall y (A(x, u) \supset U(x, u, y))$

(7) $\forall x \forall u \forall y \exists w (A(x, w) \vee (A(x, u) \supset U(x, u, y)))$

If (5) is satisfiable, then by the second half of (5), we can get (2) from the first half of (5). Conversely, if (2) is satisfiable, we can interpret $A(x, u)$ as $\forall y U(x, u, y)$.

1.1 *The schema (2) is satisfiable in a domain if and only if (6) or (7) is.*

THEOREM 2 The Löwenheim-Skolem submodel theorem. *If a quantificational schema W has a model M, then it has a countable model M' which is an elementary submodel of M (i.e. the domain D' of M' is a subset of the domain D of M, and the predicate letters get in M' the same interpretation as in M only restricted to D'). The same is true if instead of W a countable set of schemata is assumed to have a model.*

Skolem begins by considering a simple case:

(8) $\forall x \exists y U(x, y)$

Assume that (8) is satisfied in a given domain D for certain interpretation of the predicate letters. "Then, by virtue of the axiom of choice, we can imagine that for every x a uniquely determined y is chosen in such a way that $U(x, y)$ comes out true". This defines a mapping y_x of the domain D into itself. Then

(9) $\forall x U(x, y_x)$

is true for the given interpretation of the predicate letters. Let a be a particular individual and D' be the intersection of all classes X such that $a \in X$ and $y_a \in X$ if $x \in X$. Then D' is (either finite) or countable. But (9), and therefore (8), also holds true in D'.

This sketch is followed by an elaboration of the general case of a single or countable set of schemata each in Skolem normal form. The same method applies also to schemata in the prenex form, and Skolem remarks on this in [1928a] and [1941a].

The remainder of §1 of [1920a] is taken up by fairly direct results on "infinitely long expressions".

Strictly speaking, the proof deals with extended quantification theory because $=$ is implicitly included as a "relative" (predicate letter) with a fixed interpretation. The inclusion of equality does not affect the general proof but requires merely a slightly more complex notion of a model (viz. with an additional clause on $=$).

In recent years, an obvious generalization of Theorem 2 has been applied in model theory: we can begin with, instead of $\{a\}$, and arbitrary infinite subset D_1 of D, and obtain an elementary submodel of M with D_1 contained in its domain which has the same cardinality as D_1; a related result is to start with any infinite model M and a superset D_2 of D and reach a model which contains M as an elementary submodel and whose domain has the same cardinality as D_2.

In §3 of [1922b], Skolem proves a different version of Theorem 2 which avoids the axiom of choice.

THEOREM 3 The Löwenheim-Skolem approximation theorem. *If a quantificational schema is satisfied in any domain at all, it is already satisfied in a countable domain. Similarly with a countable set of schemata.*

The proof begins with a single schema in Skolem form assumed to be satisfiable:

(10) $\forall x_1 \ldots \forall x_m \exists y_1 \ldots \exists y_n U(x_1, \ldots, x_m; y_1, \ldots, y_n)$

While Skolem's proof here is correct, it is not sufficiently explicit at two points. The same ideas are elaborated and improved in §4 of [1929b] beginning with a schema in prenex form. Let us consider the two proofs together. Since the more general case of a schemata in the prenex form involves no essential difference, we confine our attention to (10).

Since (10) is satisfied in a (non-empty) domain D, we can take an arbitrary element of D and denote it by symbol 1. Then there must be certain n-tuple $y_1, \ldots, y_n$ in D such that $U(1, \ldots, 1; y_1, \ldots, y_n)$ is true (in the given model). Without loss of generality, we can denote these elements by $2, \ldots, n+1$. In other words, there must be ways of assigning truth-values to atomic schemata in $U(1, \ldots, 1; 2, \ldots, n+1)$ so that the whole schema gets the value "true". We say that (10) has solution of level 1. This does not exclude the case that some or all the objects denoted by symbols $1, 2, \ldots, n+1$ might be the same. (In the proof of 1922b, a slight complication is introduced by saying merely that some n-tuple $(y_1, \ldots, y_n)$ can be selected from $\{1, \ldots, n+1\}^n$). Let

$$c_1 = n + 1 \ , \ c_{k+1} = c_k + n(c_k^m - 1)$$

For each k, consider the conjunction K_{k+1} of all schemata

$$U(a_1, \ldots, a_m; b_1, \ldots, b_n)$$

where $(a_1, \ldots, a_m)$ is any m-tuple (say the $(t+1)$-st in a reasonable enumeration) from $\{1, \ldots, c_k\}^m$ and $(b_1, \ldots, b_n)$ is $(nt+1, \ldots, nt+n)$. A model of K_{k+1} is said to be a solution of (10) of level $k+1$. Since (10) is satisfiable, it has solutions of level k, for every k. Consider now the totality of all partial solutions of (10), i.e. the totality of sets of truth-value assignments which are solutions of (10) of level k, for some k.

For each k, there may be many (but always a finite number of) solutions. For any two different solutions L and L' of arbitrary level, write $L < L'$ if and only if $R(i, j, \ldots)$ gets "false" in L and "true" in L', where $R(i, j, \ldots)$ is the first atomic schema having different truth-values in L and L'. When $p < q$, every solution of level q must be an extension of some solution of level p. Hence, at each level k, there must be solutions which have infinitely many extensions (i.e. solutions of higher levels which extend them). (This step is missing in the proof of [1922b] which is, therefore, incomplete). Let these be $L_1^{(k)}, \ldots, L_{t_k}^{(k)}$ arranged in the order $<$ just introduced. For each k, $L_1^{(k+1)}$ must be an extension of $L_1^{(k)}$. Otherwise, since $L_1^{(k)}$ is indefinitely extendable, there must be $L_i^{(k+1)}$ which is an extension of $L_1^{(k)}$ such that $L_1^{(k+1)} < L_i^{(k+1)}$. Since $L_1^{(k+1)}$ is not an extension of $L_1^{(k)}$, $L_i^{(k+1)}$ must differ from $L_1^{(k+1)}$ in the assignments to the part K_k. But then there must be $L_j^{(k)}$ which is a part of $L_1^{(k+1)}$ such that $L_j^{(k)} < L_1^{(k)}$. This is not possible. Hence, $L_1^{(1)}, L_1^{(2)}, L_1^{(3)}, \ldots$ yield a simultaneous solution of (10), and therefore (10) is satisfiable in the domain of positive integers.

In [1928a], Skolem introduces a refutation procedure **P** and asserts its completeness:

P 1. Usual rules to turn a schema into prenex form.

P 2. Turn the schema into functional form: e.g. ;

(11) $\forall x \forall y \exists u \forall v \exists w U(x, y, u, v, w)$

into

$$U(x, y, f(x, y), v, g(x, y, v))$$

P 3. Use $0, f(0,0), g(0,0,0), \ldots$ to perform the role of $1, 2, 3, \ldots$ in the preceding proof and define solutions of level k, for each k.

P 4. Try successively for each k, whether (11) has a solution of level k.

A schema is refutable in **P** if and only if there is some k such that its prenex form has, for some k, no solution of level k.

THEOREM 4. *The refutation procedure **P** is complete; i.e. a schema W is refutable by **P** if and only if it is not satisfiable.*

The proof of this theorem is directly from Skolem's other works as quoted a moment ago. Thus, if W is satisfiable, we see from the previous arguments that W has a solution of level k, for every k, and, therefore, that W is not refutable by **P**. On the other hand, if W is not refutable by **P**, then it has a solution of level k, for every k. Hence, by the previous arguments, W is satisfiable (in fact, in a countable domain).

This differs from the famous completeness theorem only in the use of an unfamiliar refutation system in place of a familiar proof system. In fact, it is a relatively small step to establish the completeness of a proof system S (say that of Hilbert–Ackermann) by showing that if $\neg W$ is refutable in **P**, then W is provable in S.

However, Skolem's arguments in his [1928a] paper, viewed as a proof of Theorem 4, are entirely inconclusive. Specifically, the inference from the existence of solutions of every level to the satisfiablity of W is defective. In the first place, instead of beginning with the assumption that W has no model, an (informal) derivation of a contradiction from W is assumed. What confuses the issue is Skolem's free interchange of "satisfiable" and "consistent". This is explained by the fact that Skolem is in the tradition of Boole, Schröder, Löwenheim, and Korselt. According to this tradition, unlike that of Frege and Hilbert, logic is not thought of as a deductive system. From this point of view, as Professor Bernays has pointed out, satisfiability is the same as consistency (non-contradiction), since, in the domain of pure logic, no other hindrances for satisfying conditions can exist besides contradiction. Hence, for Skolem, the completeness of the refutation system **P** can only mean that if there is an informal disproof of W, then W is disprovable in **P**. In fact this completeness theorem is stated fairly explicitly by Skolem (line 14

to line 10 from the bottom, p. 198): if W is contradictory, then there is some n such that there is a contradiction at level n. This is obviously sufficient since, by definition, such a number n can then be found by the rules of the system **P**.

The structure of Skolem's proof of the above statement is as follows (the last 5 lines of p. 196 and the first 15 lines of p. 197; the last 10 lines of p. 198 and the first 3 lines of p. 199). (α) An informal derivation D of a contradiction from W is assumed. (β) The schema W is turned into functional form W_f. (γ) Instead of D, an assumed derivation D_f of a contradiction from W_f is considered, where D_f consists of substitutions and propositional combinations ("Every consequence of the axiom results from repeated and combined use of it"). (δ) By substituting 0 for all free variables in the assumed deduction D_f, a number n is deduced such that a contradiction is forthcoming at level n (and, therefore, W has no solution of level n).

Actually step (α) is left implicit because it is asserted that (i) W_f faithfully represents W. It is further assumed that (ii) all possible deductions from W_f must be sufficiently simple to satisfy (γ) and (δ). Since no formal system for codifying D and D_f are given, the whole proof is somewhat vague. However, since (i) and (ii) sound highly plausible, this approach to the completeness question of the system **P** is not without appeal. One way to avoid the vagueness is to use the arguments for proving Theorem 3. It is shown that there that if there is no model of W, then there is a contradiction at some level n. Hence, since evidently a correct informal disproof of W implies the nonexistence of models for W, we arrive at Skolem's desired conclusion. But Skolem apparently does not wish to go through the nonfinitary reasoning involved in the proof of Theorem 3 but attempts instead to eliminate transfinite arguments from metamathematics. As a result, he has to depend on the vague assumptions (i) and (ii). (By the way Skolem's assertion 1.3 described below is open to the same objection.)

At one place (p. 196, line 12 from below), Skolem states that he is trying to reduce the deduction question of logic to an arithmetical-combinatorial question. Here again, assumptions like (i) and (ii) seem to be used. In a sense, the Herbrand type of work may be viewed as a partial realization of Skolem's program, partial because Herbrand begins with a deductive system S, assumed to be adequate for all informal arguments of a logical nature. The Herbrand type of work establishes, with more complex proofs, Skolem's assertion to the extent that (a) deductions in S from W by quantifier theory theory are represented by deductions from W_f by substitutions and truth-functional steps. The interesting assertion (a) shows that if S is complete, then **P** is complete. But it does not yield Theorem 4. In fact, (a) is not relevant to the classical completeness result. Rather it proposes to give a sharper interpretation of the highly non-constructive remark that all theorems of S are valid. Herbrand does state in effect that (b) if W is valid (in a restricted sense in a countable domain), then W is provable in S [1929, C. R. vol. 188, p. 1077, Theorem II; 1930, dissertation, Chapter 5, 6.2, Theorem 2]. Classically, of course, the completeness of S follows as a corollary. However, the exact way in which Herbrand proposes to prove (b) (in particular, also 1.2 below) is infected

by a basic (correctable) error, which is crucial to his proof of (a) and is noted by several people.

The completeness theorem, mathematically, is indeed an almost immediate consequence of Skolem [1922b; 1929b]. However, the fact is, before Gödel nobody (including Skolem himself) had drawn this conclusion. Professor Gödel attributes this surprising failure to a widespread lack, at that time, of the required epistemological attitude toward metamathematics and toward nonfinitary reasoning. That Skolem used nonfinitary reasoning to improve Löwenheim's theorem does not disprove this interpretation. Pure model theory, where the concept of proof does not come in, lies on the borderline between mathematics and metamathematics, and its application to special systems with a finite number of axioms actually belongs to mathematics, at least for the most part. Skolem undoubtedly felt that one has to talk about more tangible things in a proof of completeness. Hence, his failure to appeal to the argument of [Skolem, 1922b], as well as his unsuccessful attempt to supply a syntactical argument.

The criterion of a complete system of logic was first introduced by Bolzano in *Wissenschaftslehre* [1837]. In 1928, Hilbert and Ackermann rediscovered the formulation and presented it in connection with their exact system. "Whether the axiom system is at least complete in the sense that all logical formulas, which are valid (richtig) for all domains of individuals, are in fact derivable, remains an unsolved question. It can only be stated that this axiom system is adequate to all applications" (p. 68). "All examples of formulas, which are valid (allgemeingltig) in every domain of individuals, can be proved from the axioms of the functional calculus set up by us. It is conjectured that (not only these examples but indeed) all such valid formulas are provable in the given system" (p. 80).

It is perhaps surprising that the problem of completeness, once so posed, was not solved as more or less an exercise by applying Skolem's proof of [1922b]. The essential missing step is merely (compare the notation above):

1.2 *If there is some k, K_k is contradictory (i.e. , there is no solution of level k), then the negation of (10) can be derived in S from $\neg K_k$ which is a tautology.*

Even this step was available in Herbrand's work [1929, C. R. vol. 188, p. 1076; 1930, dissertation]. But Herbrand refrained from drawing the conclusion, undoubtedly because he regarded the above concept of completeness as meaningless on account of it being highly nonconstructive.

The completeness theorem is due to Gödel [1930] and his proof is widely adopted in expositions. Gödel was not acquainted with Skolem [1922b], nor later papers of Skolem and Herbrand. Hence, it is not pointed out in Gödel's paper that the mathematical core is similar to §3 of [Skolem, 1922b]. Gödel studied Hilbert–Ackermann and [Skolem, 1920a], and then gave an attractive treatment of the whole of the whole problem of the completeness of elementary logic. In his proof, Gödel appeals to familiar arguments ("so folgt nach bekannten Schlussweisen") in a step that corresponds to the inference from the existence of solutions at each level to the existence of a sequence of solutions each containing the preceding

(p. 350). This refers to the infinity lemma; in fact a complete proof is given in Gödel's dissertation (pp. 20–22). In any case, Gödel is not interested in avoiding an appeal to the axiom of choice. The omitted step is elaborated in the exposition of Hilbert–Ackermann (second edition, [1938, p. 80]), but many may have found the elaboration not sufficiently explicit since an implicit appeal is made to the axiom of dependent choice. In contrast, Skolem is much more explicit and states in [1929b] first the principle of choice required and then proceeds to give the proof summarized above which avoids the axiom of choice.

The submodel theorem implies a certain incompleteness of an axiomatic set theory with finitely many axioms because an expression can be found in the system to enumerate the elements of the countable submodel and the enumeration, though expressible, cannot define a set in the system on account of the diagonal argument. This general idea can be applied to show, for example that the Zermelo-Fraenkel set theory has no finite axiomatization.

Using Gödel's method of arithmetization, a sharper form of the completeness theorem is obtained in Hilbert–Bernays II. This can be further modified to give the following theorem:

THEOREM 5. *If we add Con(S) (the usual formula expressing the consistency of a first order system S) to number theory as a new axiom, we can prove in the resulting system arithmetic translations of all theorems of S.*

There is a somewhat artificial question about the exact form of the formula Con(S), especially when S has an infinite number of axioms. One way to avoid this complication is to use a finite axiomatization of quantification theory with additional rules of inference. Then the systems with finitely many axioms which have interested us all receive finite axiomatization, and there is no longer room for giving queer forms of Con(S).

Herbrand sometimes speaks of his fundamental theorem as "a more precise statement of the well-known Löwenheim–Skolem theorem". Herbrand does not accept the general concepts of a domain, model, satisfiable, valid, but makes heavy use of provability in a more or less standard formal system H of quantification theory. For the present purpose, Herbrand's theorem can be stated as follows:

THEOREM 6. *(1a) If a schema W does not have a solution of level k, for some k, then W is refutable in H (i.e. $\neg W$ is provable in H). (1b) Given the number k, we can construct a refutation of W. (2a) If W is refutable in H, then there is a number k such that W has no solution of level k. (2b) There is an elementary method by which, given a refutation of any schema W, we can find a number k such that W has no solution of level k.*

This then is the constructive version of the Löwenheim–Skolem theorem; it moves essentially between two integers (a proof and a level). Of the four parts, the first three do not go beyond some familiar proofs of the completeness theorem which extend only slightly the ideas of Skolem discussed above. Part (2b) is an important result in proof theory and calls for syntactic considerations. It has been applied to obtain consistency results of mathematical theories of an elementary

sort. It contains the cut-elimination theorem commonly attributed to Gentzen.

There are certain inconclusive considerations in Skolem [1928a] and [1929b, end of §4] which have a similar flavor as this part (2b). Skolem asserts that:

1.3 *If there is, for every k, a solution of W of level k, then the theory obtained from quantification theory by adding W as an axiom is consistent.*

The argument begins with a schema W in functional form, say:

(12) $U(x, y, f(x, y), v, g(x, y, v))$

The individual constants are those built up from 0 by iterated applications of f and g. It is argued that if arrive at a contradiction, then we can replace all individual variables with 0, so that we have a schema which is contradictory but contains only individual constants of level k, for some k. Hence, (12) can have no solution of level k. It is not clear how we can turn this suggestive sketch into a convincing proof.

In §5 of [1929b], it is proved that if a schema W has a solution of level k, for each k, then the conjunction W with schemata of the form

(13) $\exists y \forall x(\neg V(y) \vee V(x))$

also has a solution of level k, for each k. Then 1.3 is applied to conclude that the addition of the law of excluded middle (13) introduces no new contradictions. It is hard to evaluate the importance of proofs and comments in this section.

There are philosophical problems connected with the interpretation of the Löwenheim–Skolem and the completeness results. As we have seen, Herbrand adheres consistently to a constructive position and declines to accept the classical interpretations of these results. Gödel has in his dissertation (though not in the published version) a characteristically lucid and balanced discussion of the foundational and methodological issues connected with the completeness result. Gödel emphasizes that the problem of completeness (unlike that of consistency) did not arise from the foundational debates and can be asked and answered meaningfully in the framework of ordinary and "naive" mathematics. Hence, no restrictions on the the methods of proof are made. Skolem's position on these issues is somewhat ambiguous. On the one hand, he discourses rather freely on arbitrary domains (e.g. in [1922b] and [1929b]) and his results dealing with arbitrary domains are of central importance in current model theory. On the other hand, in [1955c], he voices unhappiness over the uncritical use of concepts such as satisfiable, domain, valid: "What is the totality of all domains? Now it is possible by use of Löwenheim's theorem to simplify the definition of satisfiability by restricting the variable predicates occurring in the considered formula to the domain N. However, the formulation of Löwenheim's theorem itself requires either a domain in general namely by saying that if a formula is satisfied in an arbitrary domain then it is satisfied in N, or we must formulate the theorem by saying that if $\neg F$ is not provable, then F can be satisfied in N".

Skolem's critical remark contains, among other things, two observations. First "it is highly justified to ask what the starting point shall be, whether logic or set theory". It is, Skolem seems to assert, inappropriate to use set theory in studying logic and at the same time say that logic is intended to give the foundations of all mathematics. Second, Skolem seems to say that there is not one set theory but different axiom systems of set theory. "Sometimes I have had the task to write reviews of articles, where set theoretic notions and theorems are used without any explanation of what kind of set theory is meant. It is a disagreeable job to write reviews in such cases. One does not know what the author really means". If we accept completely these observations, then they would be objections against Skolem's own theorems 2 and 3 given above. Implicit in Skolem's views seems to be the suggestion that since no axiom system of set theory is categorial, then there can be no stable concept of set; that a mathematical concept is clear (or even acceptable) only if we have (at least) a characterization of it by a formal system.

A digression on Löwenheim [1915]

In Löwenheim [1915], there is a proof of Theorem 3 which is more informal and more complex than Skolem's. Confining to essentials, we may describe the proof as follows.

Roughly speaking, every quantificational schema is first transformed into one in prenex form, e.g.:

(14) $\forall x \exists y \forall z U(x, y, z)$

It is asserted that, for a suitable function y_x, (14) is equivalent to

(15) $\forall x \forall z U(x, y_x, z)$

in the sense that (14) is satisfiable in D if and only if (15) is. Obviously (14) is satisfiable if (15) is. The reason Löwenheim gives for the other direction amounts to saying that if (15) is not satisfiable in D, however we interpret the predicate letters and function symbol y_x, then (14) is not satisfiable in D. This makes an implicit appeal to the axiom of choice.

The use of "Skolem functions" seems to go back to the logicians of the Scröder school to which Löwenheim and Korselt belong. They speak of a general logical law (a distributive law) which, in modern notation, states:

$$\forall x \exists y A(x, y) \equiv \exists f \forall x A(x, f(x))$$

As Skolem comments in §3 of [1922b]: "Löwenheim's proof is unnecessarily complicated, and furthermore the complexity is rather essential to his argument, since he introduces subscripts when he expands infinite logical sums or products; in this way he obtains logical expressions with an uncountably infinite number of terms. So he must make a detour, so to speak, through the uncountable".

In fact, this whole complex preliminary does not seem necessary for Löwenheim's proof since he continues with an argument somewhat similar to Skolem [1922b; 1929b]. He introduces integers much as Skolem does and considers (in somewhat different terminology) solutions of different levels. His argument is less sophisticated than Skolem's in proving that if a schema has a solution of level k, for each k, then it is satisfiable (in the domain N of positive integers). Suppose a schema has solutions for each level k. Let E_k be the finite set of solutions of level k and E be the infinite union of these sets $E_1, E_2, \ldots$. Within E_1 there must be a solution Q_1 of level 1 which occurs as a part of infinitely many members of E since each of the infinitely many solutions in E contains one of the finitely many members of E_1 as a member. Similarly, in E_2 there must be a least one Q_2 which contains Q_1 as a part and occurs in infinitely many members of E. And so on. Hence we can take the union of $Q_1, Q_2, \ldots$ and give an interpretation of the given schema in N.

Löwenheim's argument establishes in a typical case what is today known as the "infinity lemma". He makes an implicit appeal to the axiom of dependent choice which is avoided in Skolem [1922b], as discussed above. Löwenheim's treatment also includes equality as part of the quantification theory.

With regard to the consequences of his famous theorem, Löwenheim considers an expression F which is intended to say that the universe of discourse is neither finite nor countable. The expression F contains bound predicate or class or second-order variables. Löwenheim that as the number of individuals approaches infinity, the number of possible classes classes approaches the continuum. He does not seem to envisage the possibility of expressing uncountability in a first-order language or of extending his result to second-order theories. In fact, his conclusion is rather that it is not always possible to treat second-order quantifiers in a similar way: otherwise his schema F, which, according to him, can have no finite or countable model, would have, by his theorem, a finite or a countable model.

We make a digression on the infinity lemma:

1.4 *Every infinite tree (or fan) contains an infinite path.*

Several things are implicit in the statement. In particular, it is to be understood that, from each node, only finitely many branchings are permitted. The intuitive proof of 1.4 is quite immediate, much as in Löwenheim's argument described above.

A statement not using concepts of graph theory is:

1.5 *Given a sequence of (mutually disjoint) finite nonempty sets $A_1, A_2, \ldots$ and a relation S such that $y \in A_{n+1}$ if xSy and $x \in A_n$, and for every $x_{n+1} \in A_{n+1}$ there is one (and only one) $x_n \in A_n$, $x_n S x_{n+1}$. Then there is a sequence $y_1, y_2, \ldots$ such that $y_i \in A_i$ and $y_i S y_{i+1}$ for all i.*

This can be proved by an appeal to the axiom of dependent choice:

1.6 *If B is a nonempty set and S a relation such that*

$$\forall x(x \in B \supset \exists y(y \in B \wedge xSy)$$

then there exists a sequence of elements $x_1, x_2, \ldots$ *of* B *such that* $x_n S x_{n+1}$ *for all* n.

It is well-known that 1.6 is a direct consequence of the usual axiom of choice in the form of obtaining a function from a relation R such that $\forall x \exists y R(y, x)$.

To prove 1.5 from 1.6, we delete first all those nodes which do not occur in infinitely many paths. Thus since the union A of $A_1, A_2, \ldots$ is infinite but each A_i is finite, we can make the deletion by induction and be sure that the reduced $B_1, B_2, \ldots$ all remain nonempty. Since A_1 is finite, there must be some x_1 in A_1 which bears directly or indirectly the relation S (or its iteration) to infinitely many members of A. Define

$$\begin{aligned} Px &\equiv \text{there are infinitely many members } y \text{ of } A \text{ such that } xS^n y \text{ for some } n \\ B_1 &= \{x|(x \in A_1 \wedge Px)\} \\ B_{m+1} &= \{x|(x \in A_{m+1} \wedge \exists y(y \in B_m \wedge ySx) \wedge Px)\} \end{aligned}$$

It is clear that if B_m is nonempty, then, since A_{m+1} is finite, B_{m+1} is nonempty. Let B be the union of $B_1, B_2, \ldots$ and apply 1.6 to get 1.5. In fact, we can avoid the recursive definition of B_m and define directly: $B = \{x|(x \in A \wedge Px)\}$. This is so because we can prove:

$$\forall x((x \in A_m \wedge Px) \supset \exists y(y \in A_{m+1} \wedge xSy \wedge Py))$$

Note that if we are given in advance orderings of all members of each A_i in 1.5, we can prove 1.5 without appealing to the axiom of (dependent) choice:

$$\begin{aligned} y_1 &= \text{the first member of } A_1 \text{ such that } Py_1 \\ y_{m+1} &= \text{the first member of } A_{m+1} \text{ such that } y_m S y_{m+1} \text{ and } Py_{m+1} \end{aligned}$$

By the way, there is an ambiguity in Löwenheim's assignment of truth values to atomic formulas. I have chosen to interpret him as taking all consistent assignments. The alternative interpretation of accusing him of using one fixed assignment for all except those atomic formulas containing the equality sign seems to me to involve too many difficulties.

3.1.2 Decidable classes and reduction classes

This section deals with §4 of [1919a], §1 of [1920a], [1928a; 1933c; 1935g; 1935a; 1935f], and [1936f].

A reduction class (with respect to solvability) is a class K of quantificational schemata for which there is a method by that, given any quantificational schema W, we can find W' in K such that W is satisfiable if and only if W' is.

Löwenheim 1915 contains the earliest results on both reduction classes (the class of schemata with dyadic letters only) and decidable classes (the class of schemata with monadic predicate letters only). In §4 of [1919a], Skolem gives incidentally a new proof of the decidability of monadic schemata.

In [1928a], Skolem gives two decidable classes:

2.1 *The class of schemata of the form*

$$\forall x \exists y_1 \ldots \exists y_n U(x, y_1, \ldots, y_n)$$

is decidable (i.e. there is a method by which, given any schema, we can decide whether it is satisfiable).

2.2 *The class of schemata of the Skolem form*

$$\forall x_1 \ldots \forall x_m \exists y_1 \ldots \exists y_n U(x_1, \ldots, x_m; y_1, \ldots, y_n)$$

in which each atomic part contains at least m distinct variables, is decidable.

Skolem obtains these results by applying his refutation procedure, viz. the part of Herbrand's theorem which has in recent years been employed for deciding these and other classes, largely without awareness of Skolem's early use of the same method.

2.1 is independently obtained by Ackermann [1928, Math. Annalen, vol. 100, pp. 638–649]. In [1935g], Skolem extends 2.1 and 2.2 to

2.3 *The class of schemata of the form*

$$\forall x \exists y_1, \ldots, y_n U(x, y_1, \ldots, y_n) \wedge \forall x_1 \forall x_2 \forall x_3 V(x_1, x_2, x_3)$$

where U contains only dyadic predicate letters and V is a conjunction which says that these predicates are symmetric and transitive, is decidable.

2.4 *The class of Skolem schemata, in which each atomic part contains either at least one of $y_1, \ldots, y_n$ or all of $x_1, \ldots, x_m$, is decidable.*

In [1935f], Skolem further extends 2.3 by permitting other predicate letters in U and stipulating V to be the conjunction of m^3 schemata $A_{ijh} (1 \leq i, j, h \leq m)$:

$$(R_{ij}(x_1, x_2) \equiv R_{ji}(x_1, x_2)) \wedge ((R_{ij}(x_1, x_2) \wedge R_{jh}(x_2, x_3)) \supset R_{ih}(x_1, x_3))$$

In [1933c], Skolem gives a new proof of Ramsey's theorem [1929, Proc. London Math. Soc. vol 30]; see below) and improves Ramsey's bound to the size of domain

in which a satisfiable schema, which begins with only universal quantifiers followed by a matrix containing also equality, is already satisfied.

Skolem's best known contribution to the reduction problem is his normal form (from §1 of [1920a]) quoted as Theorem 1 above. This can be restated:

2.5 *The schemata in the Skolem normal form make up a reduction class.*

In [1933d], Gödel simplifies the Skolem form to three universal quantifiers (Monatsh. Math. Phys. vol. 40, pp. 433-443). In [1935a], Skolem gives an alternative proof of this and applies the proof to give a new reduction class:

2.6 *The schemata of the form*

$$\forall x_1 \exists y_1 \dots \exists y_n U \wedge \forall x_1 \forall x_2 \exists z_1 \dots \exists z_m U_1 \wedge \forall x_1 \forall x_2 \forall x_3 V \wedge \forall x_1 \forall x_2 V_1$$

with U, U_1, V, V_1 *in suitable special forms, form a reduction class.*

In [1936f], Skolem gives several reduction classes in which each schema contains only a single predicate letter, in fact a dyadic one, as variants of a previous result by Kalmar.

Much progress has been made since Skolem worked on reduction classes. The strongest central result would seem to be the following:

2.7 *The schemata of the form* $\forall x \exists y \forall z U(x, y, z)$ *with dyadic predicate letters only, make up a reduction class; this is true even if* U *is restricted to contain only a single dyadic predicate letter plus monadic predicate letters.*

This result has various direct and indirect corollaries. By 1.1, it follows directly that schemata of the form

$$\forall x_1 \forall x_2 \forall x_3 \exists y U(x_1, x_2, x_3, y) \text{ (or } \forall x_1 \forall x_2 \forall x_3 U(x_1, x_2, x_3) \wedge \forall x_1 \exists y V(x_1, y))$$

form a reduction class. This result had been known before 2.7 was found, but the above remarks supplies a measure of unity. In fact, with the help of 2.7, it is easy to derive from other known results a general conclusion on all prefix classes (a prefix class is determined by a given string or set of strings of quantifiers in the sense that it contains all and only schemata in prenex form having the given string or a string in the given set as prefix; one could extend the next theorem to conjunction of parts each with some given string as prefix).

THEOREM 7. *Given any string of quantifiers* $Q_1 x_1 \dots Q_k x_k$*, the prefix class determined by it is a reduction class relative to satisfiablity if and only if the string contains* $\forall\exists\forall$ *or* $\forall\forall\forall\exists$ *as an order-preserving but not necessarily unbroken substring. Each prefix class that is not a reduction class is decidable and contains no*

axiom of infinity (i.e. no schemata which are satisfiable but not satisfiable in finite domains).

Thus, by 2.7 and its corollary, if the string contains $\forall\exists\forall$ or $\forall\forall\forall\exists$, then it is a reduction class. Otherwise, it either contains three or more consecutive $\forall$'s but followed by no $\exists$'s or contains at most two consecutive $\forall$'s but no other $\forall$'s. In the first case, it falls under a general case $\exists x_1 \ldots \exists x_m \forall y_1 \ldots \forall y_m$, decided by Bernays and Schönfinkel [1928, Math. Annalen, vol. 99, pp. 342–372]. In the second case, it falls under the general case $\exists x_1 \ldots \exists x_m \forall y_1 \forall y_2 \exists z_1 \ldots \exists z_n$ decided by Gödel (1933, op. cit.) and others. In both cases, the absence of axioms of infinity is also explicitly established.

There is a slight ambiguity in the above theorem. In decidable cases, the inclusion of equality often introduces substantial complications. If we exclude equality, then Theorem 7 is firmly accomplished. But if we include equality, then there seems to remain a precise and interesting open problem:

2.8 *If equality is included in the Gödel class, i.e. the class of schemata of the form*

$$\exists x_1 \ldots \exists x_m \forall y_1 \forall y_2 \exists z_1 \ldots \exists z_n U(x_1, \ldots, x_m, y_1, y_2, z_1, \ldots, z_n)$$

is it decidable? Does every satisfiable schema in the given class have a finite model?

At the end of Gödel [1933], it is stated that the proof for the case without equality can be extended to answer both questions positively. This statement agrees with Professor Gödel's present belief. Meanwhile, several people have attempted to supply the details for such an extension but have failed so far. As can be seen from Skolem's results quoted above, prefix classes are by no means the only interesting classification for the purpose of studying decision and reduction problems of elementary logic. There, however, seems to be no general clear principle to guide a more detailed study of decidable classes and undecidable classes (in particular, reduction classes).

A digression on Ramsey's theorem

In [1929] (op. cit.), Ramsey proves the following.

2.9 *Ramsey's Theorem (finite case). Given m, n, k, we can find a number N such that if we divide all unordered k-tuples of elements of any set M with at least N members, briefly $P_k(M)$, into any n disjoint sets $C_1, \ldots, C_n$, then there is always an m-membered subset T of M so that $P_k(T)$ is included in some C_i.*

Skolem's proof of this in [1933c] is simpler and gives a smaller value N. In the basic case when $k = 2$ and $n = 2$, Skolem has:

2.10 *If $N \geq 2^{2m} - 1$, then there is always a subset T with m-members such that $P_2(T) \subseteq C_1$ or $P_2(T) \subseteq C_2$.*

To illustrate applications of the theorem, consider a given finite set of N positive integers arranged in an arbitrary order. By Ramsey's theorem, we can show that there are always at least m members either all in increasing natural order or all in decreasing natural order. Thus, each arrangement divides all unordered pairs $\{x, y\}$ $1 \leq x < y \leq N$, into two sets C_1 and C_2 according to whether the x-th member in the arrangement is less than or greater than the y-th member. Hence, there must be an m-membered subset T of $\{1, \ldots, N\}$, say $\{i_1, \ldots, i_m\}$, such that $P_2(T) \subseteq C_1$ or $P_2(T) \subseteq C_2$.

In the infinite case, Ramsey's theorem is sometimes stated thus

2.11 *Ramsey's Theorem (infinite case). Let K be an infinite set, n be a positive integer, and $P_n(k) = C_1 \cup C_2, C_1 \cap C_2 = \emptyset$. Then there is an infinite subset T of K such that $P_n(T) \subseteq C_1$ or $P_n(T) \subseteq C_2$.*

The proof of 2.11 can be reduced to the special case when $n = 2$ since for $n = m + 1$, we can one-one correlate K and K^m. We prove now the case $n = 2$.

Suppose first there exist a sequence of members $x_1, x_2, \ldots$ of K together with a sequence of subsets $K_1, K_2, \ldots$ of K such that:

- $x_1 \in K, x_1 \notin K_1, K_1 \subset K$, and $\{x_1, y\} \in C_1$ for all $y \in K_1$
- $x_{n+1} \in K, x_{n+1} \notin K_{n+1}, K_{n+1} \subset K_n$, and $\{x_{n+1}, y\} \in C_1$ for all $y \in K_{n+1}$

Then $P_2(T) \subseteq C_1$, where $T = \{x_1, x_2, \ldots\}$. Obviously, $x_1, x_2, \ldots$ are all distinct, and $K_1, K_2, \ldots$ are all infinite.

Suppose now no such pair of sequences exist. This means that no matter which member of K is taken as x_1, there is a number n, such that, however $x_2, \ldots, x_n, K_1, \ldots, K_{n-1}$ are chosen to ensure that K_n is infinite, K_{n+1} is finite for any choice of x_{n+1}. Take any fixed x_1 and consider its associated K_n. Since K_n is infinite, we can find $y_1, y_2, \ldots$ and $J_1, J_2, \ldots$ to satisfy:

- $y_1 \in K_n, y_1 \notin J_1, J_1 \subset K_n$, and $\{y_1, z\} \in C_2$ for all $z \in J_1$
- $y_{m+1} \in J_m, y_{m+1} \notin J_{m+1}, J_{m+1} \subset J_m$, and $\{y_{m+1}, z\} \in C_2$ for all $z \in J_{m+1}$

Thus,take $J_1 = \{z | z \in K_n \wedge \{y_1, z\} \in C_2\}$. J_1 is infinite because otherwise we could have used y_1 as x_{n+1} and $\{z | z \in K_n \wedge \{y_1, z\} \in C_1\}$ as K_{n+1}. Similarly, for each m, given J_m, we can take any member of J_m as y_{m+1} and take $\{z | z \in J_m \wedge \{y_{m+1}, z\} \in C_2\}$ as J_{m+1}. J_{m+1} must be infinite because otherwise we could have taken y_{m+1} as x_{n+1} and $\{z | z \in J_m \wedge \{y_{m+1}, z\} \in C_1\}$ as K_{n+1}. Hence, $T = \{y_1, y_2, \ldots\}$ and $P_2(T) \subseteq C_2$.

Intuitively it is perhaps easier to think of K as well-ordered (by the axiom of choice) and in terms of an infinite tree. Begin with the first member a of K and

make two branches by $A_1 = \{u|\{a,u\} \in C_1\}$ and $B_1 = \{u|\{a,u\} \in C_2\}$. Take the first members of A_1 and B_1 as the next nodes. Repeating this process as long as we have an infinite set along each branch, we obtain an infinite tree. There must be an infinite path with all left branches or one with all right branches. Suppose there is no infinite path with all left branches, there is a node a_n with only a right branch. Beginning with a_n, we have an infinite path with only right branches, otherwise if there were any left branches anywhere, we could have attached it directly to a_n.

If K is uncountable, it must have a large cardinality in order that 2.11 is true with a subset K of the same cardinality as K. We may think of K as ordered and interpret $\{x_1, \ldots, x_n\} \in C_1$ as $\phi(x_{i_1}, \ldots, x_{i_n})$ being true, where $x_{i_1}, \ldots, x_{i_n}$ are $x_1, \ldots, x_n$ rearranged in the given order of K. Then $P_n(T)$ may be said to be indiscernible by the formula ϕ. If we think of K as the domain of a model of a system S, then we may ask whether there is a subset T of K such that, for all n, $P_n(T)$ is indiscernible by any formula $\phi(x_1, \ldots, x_n)$ of the language of S. If an uncountable cardinal ρ is such that for every K of cardinality ρ has a subset T of cardinality ρ that is indiscernible in the above sense, ρ is called a Ramsey cardinal. The postulation of such cardinals has many interesting consequences which have been studied in recent years. For example, this implies that there are only countably many constructible sets (in Gödel's sense) of integers; that all uncountable cardinals are indiscernible in the constructible universe L; and so on. It has been proved that measurable cardinals are Ramsey.

3.1.3 Decidable theories

This section deals with §3 and §4 of [1919a], [1928a], and [1930c].

§3 of [1919a] is devoted to fragments towards a decision procedure for the elementary theory of distributive lattices. As we know now, the theory as a whole is in fact undecidable (Grzegorczyk, 1950, Fund. Math.,vol. 38).

§4 of [1919a] contains the earliest employment of the method of eliminating quantifiers in decision problems. The theory S decided is equivalent to that of a power set of a given set (a special kind of Boolean algebra). The atomic formulas of S are of the forms $a > \overline{n}$ and $a < \overline{n}$ where a is a class expression obtained from variables by intersection, union, and complementation as usual, and $\overline{n}$ is some fixed non-negative integer. Formulas in general are built up by truth functions and quantifiers as usual. The proof of decidability of S proceeds by giving a procedure to transform any formula $\exists u U(u, x_1, \ldots, x_n)$ into an equivalent formula $U'(x_1, \ldots, x_n)$ and thus to perform essentially an induction to eliminate more quantifiers in succession. Skolem allows also class constants in the basic formulas and reduces in general a formula with quantifiers to one without. It is a routine step to obtain, from Skolem's procedure, the decision procedure when constants are not included or when only constants for the universal and empty classes are included.

Skolem's theory S is easily seen to be equivalent to the theory of $\subseteq$ over the

power set of a given set. For example:

$$\begin{array}{lll} a \subseteq b & \text{for} & a \cup b \leq 0 \\ a \leq 0 & \text{for} & \forall x(a \subseteq x) \\ a \leq 1 & \text{for} & \forall x(x \subseteq a \supset x \leq 0 \vee a \subseteq x) \end{array}$$

In [1928a], Skolem sketches for illustration a proof of Langford's result [1926] that the theory of open dense ordering is decidable. Instead of "decidable" Skolem uses the term "categorical". The proof proceeds by elimination of quantifiers.

Skolem [1930c] contains a sketch of the decision procedure for the elementary theory of multiplication of natural numbers. A more elegant proof of this result, using Skolem's result on the theory S mentioned above, can be extracted from Mostowski's paper on direct products [1952, JSL]. Instead of providing a procedure by which quantifiers are eliminated in general situation, Skolem gives five "typical examples" and asserts "I believe that they are already sufficient to convince oneself that we are in a position to handle all cases". Skolem's consideration generalizes to other domains, e.g. integers over an arbitrary algebraic number field.

The paper [1930c] begins with a decision procedure by elimination of quantifiers for the following theory. A formula is obtained from atomic formulas $s = t$ and $s < t$ by the usual logical operations. A term is defined inductively: (1) a variable (ranging over all integers) is a term; (2) a term multiplied by a rational number is again a term; (3) the sum of two terms is a term; (4) if s is a term, so is $[s]$, the integral part of s. The decision result for this theory is applied both to obtain the result on multiplication just described and to construct an example to illustrate the relativity of the concepts of well-ordering and natural number. The construction gives an ordered set M of objects and a set B of predicates (propositional functions) over M which includes $<$ and is closed under the five logical operations. It is shown that, for every predicate in B, well-ordering holds, but M is no longer well-ordered when suitable new predicates are used.

3.2 Axiomatic Set Theory

3.2.1 Definite properties and finite axiomatization

This section deals with [1922b; 1929b; 1930b; 1941a], and [1958d]. Related publications include Zermelo [1908] (Math. Annalen, vol. 65, pp. 261–281), Fraenkel [1921] (abstract, Jahresbericht Deutsch. Math.-Ver., vol. 30); [1922] (paper on axiom of choice) and [1925] (Math. Zeitsch. vol. 22, pp. 250–273); von Neumann [1925] (Crelle's vol. 154) and [1928] (Fund. Math., vol. 16); Bernays [1937] (JSL, vol. 2).

Contemporary formulation of set theory usually includes the following axioms: (1) Extensionality; (2) separation; (3) power set; (4) union; (5) infinity; (6) replacement; (7) regularity; (8) choice. Often the axiom of choice is treated separately, sometimes the axiom of regularity is not included. Of these axioms, (1) to (5)

and (8) are included in Zermelo 1908, except that the axiom of separation is given a somewhat informal formulation. The current formulation of (2) and replacement appeared for the first time in Skolem [1922b]. The axiom of regularity was first suggested by Mirimanoff [1917] (L'enseignement math., vol. 19) and Skolem [1922b, §6] and adopted by von Neumann [1925] and Zermelo [1930].

It is true that in 1921 Fraenkel voices dissatisfaction with Zermelo's notion of definite property and remarks on the unavailability of the set $T = \{\omega, P\omega, \ldots\}$, giving an informal statement of the axiom of replacement. However, Fraenkel makes the first serious attempt to explicate Zermelo's only in [1922] and [1925], and unsuccessfully. As shown in von Neumann [1928], by Fraenkel's concept of definite property, 3.2 below is a consequence of 3.1 below so that he does not have a version of the axiom of replacement which goes beyond the axiom of separation. In fact, an informal version of the axiom of replacement was given by Cantor in [1899] (Gesammelte Abhandlungen, p. 444). When we consider the principal difference between Zermelo's original system and current formal system of set theory, it seems more appropriate to refer to the latter as Zermelo–Skolem rather than Zermelo-Fraenkel (an application introduced in Zermelo [1930]). However, following custom, we shall continue to speak of ZF.

The axiom of separation is stated in Zermelo [1908] as:

3.1 Axiom of separation *Whenever the class proposition $E(x)$ is definite for all elements of a set y, y possesses a subset y_E containing as elements precisely those elements x of y for which $E(x)$ is true.*

"A question or assertion E is said to be *definite* if the fundamental relations of the domain, by means of the axioms and the universally valid laws of logic, determine without arbitrariness whether it holds or not. Likewise, a class proposition $E(x)$, in which the variable term x ranges over all individuals of a class K, is said to be definite if it is definite for *each single* individual x of the class K".

In order to prove the independence of the axiom of choice (using Urelements), Fraenkel finds it necessary to render more explicit the concept of a definite property. Similarly, Skolem finds it necessary to do so in order to apply the Löwenheim–Skolem Theorem.

In [1922b], Skolem remarks on the obvious lack of precision of Zermelo's notion and goes on to list the five familiar logical operations (conjunction, disjunction, negation, universal and existential quantifications). "By a definite proposition we now means a finite expression constructed from elementary propositions of the form $a \in b$ or $a = b$ by means of the five operations mentioned. This is a completely clear notion and one that is sufficiently comprehensive to permit us to carry out all ordinary set-theoretic proofs. I therefore adopt this conception as a basis here".

After thus making precise the axiom of separation, Skolem proceeds in §4 to show that $T = \{\omega, P\omega, P^2\omega, \ldots\}$ cannot be shown to be a set in Zermelo's system (P is the power set operation). The proof amounts to arguing that the union of

T forms a model of Zermelo's set theory. Skolem states the axiom of replacement thus:

3.2 Axiom of replacement *Let U be a definite proposition that holds for certain pairs $\langle a, b\rangle$ in the domain B; assume, further, that for every a there exists at most one b such that U is true. Then, as a ranges over the elements of a set M_a, b ranges over all elements of a set M_b.*

As we know, 3.1 follows from 3.2. If we require in 3.2 that U be a one-one correspondence, then 3.1 will not follow.

In [1929], Zermelo, not knowing Skolem's work, objects to Fraenkel's concept on the ground that it uses numbers and then proposes an elaborate definition of definite properties. In [1930b], Skolem observes that one of Zermelo's requirements is rather obscure, and, if it is made precise in the natural way, Zermelo's definition agrees with Skolem's own. In Zermelo [1930], the exchange is mentioned, but Zermelo continues to use an informal concept of definite properties and states that he reserves a detailed discussion of the exchange for the future.

It remains to consider Fraenkel's proposals and von Neumann's modification thereof. In [1922], Fraenkel defines a definite property of x as $f(x) \in g(x)$ or $f(x) \notin g(x)$, where $f(x)$ and $g(x)$ are formed from x and further given objects by repeated applications of the operations of pairing, power set, union, separation, and choice. This definition is modified in 1925. By the way, the axiom of replacement is not included by Fraenkel in either 1922 or 1925.

3.3 Fraenkel [1925] *A definite property of x is of the form $f(x) \circ g(x)$, where $\circ$ is $=$, $\neq$, $\in$ or $\notin$ and $f(x)$ and $g(x)$ are functions of x. A function of x is obtained as follows:*

- *A constant or the variable x is a function of x*
- *Power set Px and union Ux are functions of x*
- *If $f(x)$ and $g(x)$ are functions of x, so are $\{f(x), g(x)\}$ and $f(g(x))$*
- *If $f(x, y)$ and $g(x, y)$ are functions of x and u is a given set, so is $h(x) = \{y|(y \in U \wedge f(x, y) \in g(x, y)\}$*

Using this definition, von Neumann [1928] shows:

3.4 *For every set u and every function f, there is a function g, such that $g(f, u) = \{y|\exists x(x \in u \wedge y = f(x))\}$.*

Further, Skolem shows in [1929b, §1]:

3.5 *Every definite property according to 3.3 is also a definite property in Skolem's sense.*

Von Neumann adds to 3.3 the following clause

- *If, for every* x, $\{y|f(x,y) \circ g(x,y)\}$ *is a set, then* $h(x) = \{y|f(x,y) \circ g(x,y)\}$ *is a function of* x.

It is easy to see that even after the modification, every definite property by 3.3 is definite in Skolem's sense. On the other hand, it seems likely that there are definite properties in Skolem's sense which go beyond the definition of Fraenkel and von Neumann.

In [1925], von Neumann introduces a systematic treatment of functions by taking function as an undefined notion and distinguishing between functions and arguments as two kinds of object. This, besides providing an alternative to Skolem's analysis of definite properties, renders it possible to use just a finite number of axioms. Skolem 1941a gives a translation of von Neumann's axioms directly into the notation of quantification theory by adding four primitive predicates. A congenial thorough reformulation of von Neumann's system is given in Bernays [1937] (and later continuations) and widely used in the literature.

In [1958d], Skolem undertakes to reduce axiom schemata to simple axioms. The result is rather similar to Bernays [1937], except that, instead of defining finite sequences, two primitive concepts (concatenation and "equal length") are introduced to deal with them directly. By using also variables over predicates, a two-sorted theory is obtained, and then it is observed that, by adding two new monadic predicates, we can turn the system into one in the usual quantification theory. It seems fair to say that Skolem [1958d] does not go essentially beyond Bernays [1937]. By the way, Skolem [1961a] does not add much to [1958d].

3.3 Skolem's paradox and impredicative definitions

It is first remarked in Skolem [1922b]: "If Zermelo's axiom system, when made precise, is consistent, it must be possible to introduce an infinite sequence of symbols $1, 2, 3, \ldots$ in such a way that they form a domain B in which all of Zermelo's axioms hold provided these symbols are suitably grouped into pairs of the form $a \in b$". Since the intended model of Zermelo's system is uncountable, this consequence of the Löwenheim-Skolem theorem has been known as "Skolem's paradox". In concluding [1922b], Skolem remarks: "The most important result above is that set-theoretic notions are relative. I had already communicated it orally to F. Bernstein in Göttingen in the winter of 1915–16. ... I believed it was so clear that axiomatization in terms of sets was not a satisfactory ultimate foundation of mathematics that mathematicians would, for the most part, not be very much concerned with it. But in recent times I have seen to my surprise that so many mathematicians think that these axioms of set theory provide the ideal foundation for mathematics; therefore it seemed to me that the time had come to publish a critique".

In §3 of [1922b], Skolem explains how, e.g. $P\omega$ in a countable model B can be uncountable "in the sense of the axiomatization". This merely means that *within* B there occurs no one-to-one mapping f of $(P\omega)_B$ onto ω. Nevertheless, $(P\omega)_B$ is countable by a collection of pairs which is not a set in the model B. It is clear

that one can define subsets of ω which are not in $(P\omega)_B$, by referring to B, e.g. $n \in D$ if and only if n does not belong to the nth member of B. The notions "finite", "infinite", "Dedekind chain", and so forth are also relative in the same sense. "Within a suitable axiomatic basis, therefore, the theorems of set theory can be made to hold in a merely *verbal* sense, on the assumption, of course, that the axiomatization is consistent; but this rests merely upon the fact that the use of the *word* set has been regulated in a suitable way". Similar remarks are made in §7 of [1929b] with special emphasis on Cantor's second number class: "It is likely that already *in* the second number class there occurs so great a number α that in order to characterize α, a logical expression of the level-number α must be employed. In any case, it is impossible to derive the second number class from simpler concepts. One recognizes here again, as with the earlier review of the Löwenheim Theorem, that there is no possibility of introducing something absolutely uncountable except by means of a pure dogma". Another elaboration of the "relativity" of set-theoretic concepts is given in §2 of [1941a].

While discussing "relativism", Skolem also considers various related questions. (1) Direct construction of countable models; (2) independence; (3) consistency; (4) categoricity; (5) impredicative definitions. And beginning with §7 of [1929b], Skolem brings in also the question of nonstandard models for arithmetic.

In §2 of [1929b], Skolem gives an axiom system of set theory in which operations for introducing new sets from given sets are explicitly given. An enumeration of all set symbols is given and these are put in $\in$-pairs and $\notin$-pairs by the axioms. It is then argued that these determine a countable model if the system is consistent. Appeal to the Löwenheim-Skolem Theorem is avoided. It is left open whether the system is as strong as ZF.

In §6 of [1922b], Skolem, while discussing the categoricity of axioms of set theory, remarks that it would be of great interest if we could extend a given model B of axioms of set theory to a new model containing a subset of ω not in B. Skolem continues to say "but this would probably be very difficult" and adds a long footnote. "Since Zermelo's axioms do not uniquely determine the domain B, it is very improbable that all cardinality problems are decidable by means of these axioms. For example, it is quite probable that what is called the continuum problem, namely whether $2^{\aleph_0} >$ or $= \aleph_1$, is not solvable on this basis; nothing need to be decided about it. The situation may be exactly the same as in the following case: an unspecified commutative field is given, and one asks whether it contains an element x such that $x^2 = 2$. This is just not determined, on account of the non-uniqueness of the domain".

In §7 of [1922b], Skolem suggests that independence and consistency results in set theory are to be studied by means of countable models: "If we do not again want to take axioms for domains as basis (and so on ad infinitum), I see no other way out than to pass on to considerations such as those employed above in the proof of Löwenheim's theorem, in which the finite integer is taken as basic". Skolem proceeds to side up with Poincaré against Hilbert to argue that expressions behave like integers and that induction is necessary for syntactic proofs of consistency.

In §6 of [1922b], Skolem considers the question whether ZF is categorical. He considers the possibility of sets which are not well-founded and speaks of $\in$-sequences:

$$\ldots \in x_2 \in x_1 \in x$$

"One now sees almost immediately that those sets in a domain B, for which the $\in$-sequences necessarily terminate, must form a part B' of B in which the axioms still hold". Skolem does not prove this in detail; in fact the familiar proof in terms of ranks goes back to a paper by von Neumann in [1929]. If therefore, Skolem remarks, we have a model which is not well-founded, we obtain immediately two non-isomorphic models. Suppose now a given model B is well-founded. Skolem goes on to outline a description of a new model B' by adding a new object a which is at the end of each branch of the "membership tree" of any set. The outline is not without ambiguity, but it seems that, however we interpret it, we would have some difficulty: it would seem that either a belongs to the empty set and extensionality fails or a is the new empty set and the new model is isomorphic to the old. Neither seems to establish the non-categorical character of the axioms. In §3 of [1941a], Skolem returns to the question of categoricity and credits von Neumann [1925] with remarking that his axioms are not categorical and that probably no axiom system of set theory can be categorical. Then Skolem quotes Gödel's incompleteness results which definitely settles this question.

In §5 of [1922b], impredicative definitions are singled out as the central difficulty in the problem of proving the consistency of set theory. In §3 of [1929b], two somewhat artificial and fragmentary axiom systems are given in which impredicative requirements are included. Models for both are given, the second with the law of excluded middle abolished. More elaborate examples are considered in §6, which concludes with the suggestion that similar considerations may yield a proof of consistency of the impredicative requirements in classical analysis.

Some of the general consequences Skolem draws from the existence of countable models seem unjustified. For example, that ZF, if consistent, has a countable model does not in itself exclude the possibility of our having an objective conception of a standard or intended model of ZF. While it is an arguable point whether there is an absolutely uncountable set, Löwenheim's theorem does not in itself justify Skolem's conclusion that it can only come from a pure dogma. Skolem seems also to overlook the distinction between relativity to an axiom system and to a given (countable) model thereof. While $(P\omega)_B$ in a countable model B of ZF is (absolutely) countable, it does not follow from this that $P\omega$ is uncountable only relative to ZF, unless the latter proposition is simply taken to mean the former. In short, the Löwenheim Theorem seems quite neutral with regard to the philosophical issue whether sets are real and in what sense they are real. Skolem has probably other subjective reasons for distrusting sets. In [1958a], Skolem proposes to present "a relativist conception of the fundamental notions of mathematics. It seems to me that this is clearer than the absolutist and Platonist conception which dominates classical mathematics". A proposition is true if it is formally proved

in a formal system. In simple cases such as recursive arithmetic, the intuition of induction is sufficient to guide the construction of the formal system and to assure consistency. "In the other case, one may accept an opportunist view. My point of view is then that one must use formal systems for the development of mathematical ideas". After listing diverse formal systems of set theory, Skolem asserts: "I do not understand why most mathematicians and logicians do not appear to be satisfied with this notion of set as defined by a formal system, but contrariwise speak of the limitations of the axiomatic method. Of course this notion of a set has a relative character, because it depends on the chosen formal system. But if this system is chosen suitably, we can nonetheless develop mathematics on the basis". Clearly these views of Skolem go beyond mathematics proper and there are many philosophical ramifications which cannot be discussed in this context.

Leaving aside Skolem's scepticism about sets, his mathematical ideas have found various important applications. Thus, Skolem functions play a central role in Gödel's proof of the relative consistency of the continuum hypothesis. Paul J. Cohen's independence proofs, at least in their original form and in their psychological origin, use countable models and Skolem functions in an essential way. The independence of the continuum hypothesis makes strong use of the availability of many subsets of ω lying outside a given countable model of ZF. In fact, using Cohen's method, it is now a common trick to "collapse cardinals" by introducing new sets which one-one correlate cardinals that were originally non-equivalent.

3.3.1 Nonstandard models of arithmetic

In §3 of [1922b], Skolem mentions in passing that the number sequence ω defined as the intersection of all sets having the same (inductive) property would be different in different models of ZF. In §7 of [1929b], it is pointed out that to characterize the number sequence amounts to requiring the validity of mathematical induction, which involves the concept of "propositional function" or "set". Hence, what the number sequence is depends on the axioms for the underlying sets. It is for the first time suggested in print (for an unpublished anticipation, compare Dedekind's letter, JSL, vol. 22, [1957, p. 150]) that given any set of theorems on natural numbers, we can find a nonstandard model in which these theorems are true. A fragmentary example is given to show that a certain set of polynomials satisfies a number of theorems on integers. In §2 of [1930c], this same set is shown to satisfy more properties including closure with respect to the five logical operations.

Finally in [1933d] (also [1934b; 1941a; 1955d]), Skolem proves a general theorem on all formulas of elementary number theory P, i.e. those obtained in the usual manner from: (1) variables ranging over natural numbers; (2) $+$ (addition) and $\cdot$ (multiplication); (3) $>$ and $=$; (4) operations of the propositional calculus; (5) quantifiers with number variables. (More complicated recursive and indeed arithmetic functions can be defined in P; alternatively, they can be added at the beginning without affecting Skolem's proof).

THEOREM 8. *There exists a system N^* of things, for which two operations* $+$

and $\cdot$, and two relations $=$ and $<$ are defined, such that N^* is not isomorphic to the system N of natural numbers, but, nevertheless, all sentences of P which are true of N are also true of N^*.

4.1 Corollary *No axiom system using only the notation of P (i.e. using only concepts of elementary number theory) can determine uniquely the structure of the sequence of natural numbers.*

The Theorem is stronger than 4.1 because it applies also when the given set of true sentences is not axiomatizable (i.e. not a recursively enumerable set). 4.1 also follows from Gödel's incompleteness theorem in a familiar manner.

Skolem often speaks of this theorem in terms of axiom systems. He may sometimes have in mind a broader sense of axiom system (so that the axioms need not form a recursive set, but could be any set of sentences). In any case, his proof does not depend on one's beginning with an axiom system in the currently accepted effective sense.

Skolem's proof uses Skolem functions to get rid of quantifiers and a combinatorial lemma which is quite general:

THEOREM 9. *Let $f_1(x), f_2(x), \ldots$ be an infinite sequence of functions of natural numbers. Then there exists a monotone increasing function $g(x)$ such that, for any pair f_i and f_j one of the three following cases holds for almost all x (i. e., for all but a finite number of them):*

$$f_i(g(x)) < f_j(g(x)) \ , \ f_i(g(x)) = f_j(g(x)) \ , \ f_i(g(x)) > f_j(g(x))$$

Let $f_i \sim f_j$ if $f_i(g(x)) = f_j(g(x))$ for almost all x. In other words, if C is the set of all cofinite subsets of ω.

4.2 *$f_i \sim f_j$ if and only if $\{t | f_i(g(t)) = f_j(g(t))\} \in C$.*

For the proof of theorem 8, Skolem uses theorem 9 and takes the set of equivalence classes (by 4.2) of all definable functions of one variable as $N^* = \{f_1, f_2, \ldots\}/\sim$.

There are various ways of generalizing Skolem's argument for proving theorem 8. Roughly speaking, it provides a general method for getting, from a given model M of a theory T a new and non-isomorphic model M^* of T. To generalize the proof of theorem 8 directly, we have to, among other things, make sure that the negation of each atomic formula can be expressed as a positive truth function of atomic formulas.

The current generalization of Skolem's is to ultrapowers and ultraproducts. There is a body of literature on these and related topics.

In the general case, Skolem's method of constructing non-standard models is highly non-constructive, and we do not know enough about these models to say anything specific such as applying them to prove independence. In [1955d], Skolem

gives some examples for fragments of number theory and proposes a conjecture on the independence of "$\sqrt{2}$ is irrational" from suitable axioms. The question is answered by Shepherdson ([1964], Bull. Polish Acad. Sci., vol. 12). Shepherdson and others have a number of papers in this area.

3.3.2 Approximations to the full comprehension axiom

Skolem has five papers dealing with approximations to the unrestricted comprehension axiom:

5.1 $\exists y \forall x (x \in y \equiv A(x, x_1, \ldots, x_n))$ *for every formula* A *of set theory.*

In [1952b], it is shown that even if we delete negation and use only positive quantification we can still get a contradiction from 5.1. Let $b = \{x|(x \in x \supset B)\}$, for any statement B. By 5.1,

5.2 $b \in b \equiv (b \in b \supset B)$

Assume $b \in b$, we get B by modus ponens. Hence, by the deduction theorem, $b \in b \supset B$. By 5.2 we get $b \in b$, and, hence, also B. Hence, any statement B can be derived from 5.1. This remark of Skolem's seems to be essentially the same as "Curry's paradox" of [1942] (JSL, vol. 7, p. 115).

In [1957e], Skolem begins to use many-valued logics. The general problem is, for each many-valued logic L, to find a model in the sense of L for the language of set theory so that 5.1 becomes true, i. e.:

5.3 *For all* $x_1, \ldots, x_n$ *in the model, there is a member* y *of the model such that for every* x *in the model, the truth-value of* $x \in y$ *is the same as that of* $A(x, x_1, \ldots, x_n)$.

The first remark of [1957e] is that we cannot obtain a model to satisfy 5.3 if L is Łukasiewicz's 3-valued logic. A contradiction can be got by considering $a = \{x|(x \in x \supset (x \in x \supset \neg x \in x))\}$. Then it is shown that by using a suitable infinite-valued logic, 5.3 can be satisfied for those formulas A which contain no quantifiers. Stronger results along this line have since been obtained by C. C. Chang and J. E. Fenstad.

In [1960b], a certain 3-valued logic is proposed for which it is shown that 5.3 is satisfiable for quantifier-free formulas A.

In [1960c], it is shown that 5.1 has a model M even in the 2-valued logic if A is built up from atomic formulas by $\wedge$ and $\vee$ only. In [1963c], the same model M is shown to satisfy 5.1 when A is a formula in prenex form followed by a matrix not containing negation. Next, a model is found for 5.1 when A contains negation but no quantifiers, provided no atomic formulas of the form $u \in u$ occurs. Then an improved proof of the main result of [1960b] is given. The paper concludes with countable (2-valued) models for several fragments of current set theory. "I have

not yet had the opportunity to extend these investigations to systems of axioms, where the "Aussonderungsaxiom" or cases of it are taken into account, but I don't think there will be any essential difficulties".

We also mention incidentally the two remarks of Skolem [1957b]. The first remark gives a formulation of the axiom of infinity that explicitly gives a defining predicate for the infinite set.

5.4 *$F(a,b)$ for $a \in b \vee \neg\{a\} \in b$*

5.5 *$I(a)$ is the conjunction of* (1) $0 \in a$; (2) $(E!y)F(y,a)$; (3) $\forall z(z \in a \supset \exists u F(u,z))$; (4) $\forall u((0 \in u \supseteq a \wedge (\iota y)F(y,a) = (\iota y)F(y,u)) \supset u = a)$

5.6 The axiom of infinity $\exists y \forall x(x \in y \equiv I(x))$

It is not clear that this is in any way superior to the formulations of the same form based on familiar devices of converting inductive definitions to explicit ones, e.g.

$$\exists y \forall x(x \in y \equiv \forall z((0 \in z \wedge \forall u(u \in z \supset \{u\} \in z)) \supset x \in z)$$

The second remark deals with ordered n-tuples as sets. Starting with $\langle a,b\rangle = \{a,\{a,b\}\}$, the familiar definition $\langle a,b,c\rangle = \langle\langle a,b\rangle,c\rangle$ is objected to on the ground that when a, b, c are of the same type, members of $\langle a,b,c\rangle$ are not of the same type. Two proposals are made.

5.7

$$\begin{aligned} \langle a,b,c\rangle &= \langle\langle a,b\rangle,\langle a,c\rangle\rangle \\ \langle a_1,\ldots,a_{n+1}\rangle &= \langle\langle a_1,a_{n+1}\rangle,\ldots,\langle a_n,a_{n+1}\rangle\rangle \end{aligned}$$

5.8

$$\begin{aligned} \langle a,b,c\rangle &= \{\langle a,b\rangle,\langle a,c\rangle,\langle b,c\rangle\} \\ \langle a,b,c,d\rangle &= \{\langle a,b,c\rangle,\langle a,b,d\rangle,\langle a,c,d\rangle,\langle b,c,d\rangle\} \end{aligned}$$

In conclusion, it is emphasized that "it is still a problem now how the ordered n-tuple can be defined in the most suitable way".

3.4 Algebraic Aspects

3.4.1 Lattices and related matters

In Skolem [1913b], distributive lattices are studied on the basis of partial orderings. In Skolem [1919a], certain relations between distributive and general lattices are considered. Results in these two papers are summarized and strengthened in Skolem [1936g]. On the historical question of Skolem's place as one of the originators of lattice theory, there are debates which we shall not discuss. Skolem [1958b] contains an exposition of lattice-theoretic interpretations of intuitionist logic. The isolated paper Skolem [1931d] (a shorter version has been published, Skolem [1931b]) is derived from work done at about the same time as Skolem [1913b].

In Skolem [1913b], a domain with a reflexive and transitive relation $\leq$ is considered, for which it is assumed that for any elements a, b, there are elements c, d such that, for all e

$$\begin{aligned} (e \leq c) &\equiv (e \leq a \wedge e \leq b) \\ (d \leq e) &\equiv (a \leq e \wedge b \leq e) \end{aligned}$$

We put then $c = a \cap b, d = a \cup b$. Taking the domain of integers, divisibility may be construed as $\leq$, greatest common divisor and least common multiple as $a \cap b$ and $a \cup b$.

Some calculations on free distributive lattices with n generators are made in Skolem [1913b]. These are completed by a general theorem in Skolem [1936g].

6.1 *The number of elements of the free distributive lattice generated by n symbols is one less than the number of sets of subsets of* $\{1, \ldots, n\}$ *which are "maximal" in them. A subset* K *of* $\{1, \ldots, n\}$ *is maximal in a set* S *if* $K \in S$ *and no superset of* K *belongs to* S.

Skolem [1913b] then considers what has been known as linear lattices. In addition, a process is introduced by which a product of two given lattices G and H is defined as a new lattice $G \circ H$ consisting of pairs $\langle g, h \rangle$ with g in G and h in H.

Skolem [1919a] deals also with what is known as pseudo-complemented lattices. In fact, Skolem considers two kinds of lattices A and B: In a lattice of type A, for any a and b, there is a maximum x denoted by b/a, such that $ax \leq b$; in one of type B, there is always a minimum y, denoted by $a - b$, such that $a \leq b + y$.

6.2 *Every lattice of type A or B is distributive.*

6.3 *Every finite distributive lattice is of types A and B.*

Various other results are then proved about lattices of types A and B.

In Skolem [1936g], the following additional theorems are proved.

6.4 *If L is a finite lattice of types A and B, and 0 has only one neighbor (a say, such that for no x, $0 < x < a$), then V is a linear lattice.*

6.5 *If V is a finite lattice of types A and B, and $a_1, \ldots, a_n$ are the neighbors of 0, then V is isomorphic to $V_1 \circ V'$ where*

$$\begin{aligned} V_1 &= \{x \in V | (a_1 \cup \ldots \cup a_n) \cap x \leq a_i\}, i = 1, \ldots, n \\ V' &= \{y \in V | (a_1 \cup \ldots \cup a_n) \cap y \leq a_2 \cup \ldots \cup a_n\} \end{aligned}$$

Moreover, V is also isomorphic to $V_1 \circ \ldots \circ V_n$, where each V_i is linear.

The interpretation of intuitionistic logic by lattices and topologies was studied in the 1930's. In [1958b], Skolem returns to use lattices of type A to prove two known results.

6.6 *Let H be intuitionistic propositional calculus. Let F^* be obtained from F when we replace $\wedge, \vee, a \supset b, \neg a$ by $\cap, \cup, a/b, a/0$. Then F is a theorem of H if and only if F^* is a theorem of a lattice of type A.*

6.7 *If $A \vee B$ is a theorem of H, then A or B is.*

Skolem [1931d] deals with symmetric general solutions of Boolean equations. Let

(16) $f\{a_l, x_k\} = 0$

be a Boolean equation with (possibly infinitely many) variables x_k and undetermined coefficients a_l. A general solution

(17) $x_i = g_i\{a_l, u_k\}$

of (16) is said to be symmetric relative to a group G of permutations of a_l and x_i which transforms (16) into itself, if for every S in G there is a permutation T of u_i such that the system (17) of equations transforms into itself when S and T are applied together. The solution is said to be normal if further $x_i = g_i\{a_l, x_k\}$ for all solutions x_i of (16) and the corresponding T permutes u_i in the same way as S permutes x_i.

6.8 *An equation $f\{x_k\}$ with constant coefficients has symmetric general (and also normal) solutions relative to G if and only if it remains solvable when all variables variables x_i are substituted for each other within each transitivity system into which G divides the variables.*

In the general case (16), let $\Sigma_A(Af_A(x_k))$ be the Boolean development of $f\{a_l, x_k)\}$ according to a_l and G_A be the subgroup of G which transforms A into itself. Let ΣB be the union of those constituents A of ΣA for which $f_A\{x_k\}$ has no symmetric general solution relative to G_A.

6.9 *The equation (16) has symmetric general (also normal) solutions if and only if* (a) *each* $f_A\{x_k\} = 0$ *(an equation with constant coefficients) has symmetric general solutions relative to* G_A*; and* (b) $\Sigma B = 0$.

3.4.2 A note on classical propositional calculus

We digress to solve an open question raised in Skolem [1951a]. This paper gives simpler independence proofs for the following system for Hilbert and Ackermann.

A1. $\neg(a \vee a) \vee a$

A2. $\neg a \vee (a \vee b)$

A3. $\neg(a \vee b) \vee (b \vee a)$

A4. $\neg(\neg a \vee b) \vee (\neg(c \vee a) \vee (c \vee b))$

R. If a and $\neg a \vee b$, then b

T. (A theorem). $\neg a \vee a$

The main results are an improvement from 4 to 3 truth values in the proofs of the independence of **A2** and **A3** from **A1**, **A3** (resp. **A2**), **A4**, **R**, and **T**. Since **A1** was originally proved independent by 3 values, **A4** is the only one for which no proof with 3 values is known. Skolem leaves this as an open problem: "As to the independence proof of the fourth axiom I have not succeeded in carrying it through with only three truth values. On the other hand, I have not been able to show that the proof is impossible if we use only 3 truth values". (p. 23). The following arguments seem to clarify the situation.

On the one hand, the following simple table would work with designated value $D = \{0\}$ and non-designated values $N = \{1, 2\}$.

p	000	111	222
q	012	012	012
$\neg p$	222	111	000
$p \vee q$	000	000	002

It is easy to verify that **A1**, **A2**, **A3**, **T** always get value 0, since we cannot obtain enough 2's without getting 0 by negation. **A4** is not satisfied if we put $a = 1, b = c = 2 : \neg(1 \vee 2) \vee (\neg(2 \vee 1) \vee (2 \vee 2)) = \neg 0 \vee (\neg 0 \vee 2) = 2 \vee (2 \vee 2) = 2$. The verification of **R** is slightly complicated. We wish to show that $v(b)$ is *always* 0 if $v(a)$ and $v(\neg a \vee b)$ are always 0. Since $v(a)$ is always 0, we must have also

$v(\neg a \vee b) = 2 \vee v(b)$ always 0. We wish to conclude that $v(b)$ is then always 0. Certainly $v(b) \neq 2$, otherwise $2 \vee v(b) = 2 \neq 0$. Consider the case that $v(b) = 1$ for suitable assignments of values to the components of b. If b is atomic, this is impossible because $v(b)$ could also be 2 then. If b is $c \vee d$, then it is impossible that $v(b) = 1$, by the table. Suppose now b is $\neg c$, if $v(b) = 1$, then $v(c) = 1$ and we can repeat the argument with c. Hence $v(b)$ always get the value 0.

On the other hand, most independence proofs satisfy a stronger requirement than **R**:

L1. If $v(a) \in D, v(b) \in N$, then $v(\neg a \vee b) \in N$; if $v(a) \in D, v(\neg a \vee b) \in D$, then $v(b) \in D$.

We shall show that there is no 3-valued logic satisfying **L1** which can establish the independence of **A4**.

Since **A1**, **A2**, **A3** and **T** must always get designated values, we have, using **L1** in most cases:

L2. If $v(a \vee b) \in D$, then $v(a) \in D$; therefore, if $v(a) \in N$, then $v(a \vee a) \in N$. By **A1**.

L3. $v(a \vee b) \in D$ if and only if $v(b \vee a) \in D$. By **A3**.

L4. If $v(a) \in D$, then $v(a \vee b) \in D$ and $v(b \vee a) \in D$. By **A2** and **L3**.

L5. $v(\neg a \vee a) \in D$; $v(a \vee \neg a) \in D$. By **T** and **L3**.

L6. If $v(a) = n \in N$, then $v(\neg a) \neq n$. Otherwise, $v(\neg a \vee a) = n \vee n$. By **L2**, $n \vee n \in N$, contradicting **L5**.

L7. If $v(a) = d \in D$, then $v(\neg a) \in N$. Otherwise, we can have $v(b) = n \in N$, $v(a) = d$ and $v(\neg a \vee b) = \neg d \vee n = d' \vee n$. By **L4**, $d' \vee n \in D$, contradicting **L1**.

We distinguish now between two cases.

Case I. *There are two designated values, say* $D = \{0, 1\}, N = \{2\}$.

In this case, **L4**, **L2**, **L6**, **L7** narrow down the possible truth tables to those of the following form:

p	q	$p \vee q$	$\neg p$
d	d	d	2
d	2	d	
2	d	d	d
2	2	2	

However, no such table will establish the independence of **A4**. Thus, in order that the value of **A4** be 2, we must have $v(c) = v(b) = 2$, but then $v(\neg(c \vee a)) = d$ if $v(a) = 2$, and $(v(\neg(\neg a \vee b)) = d$ if $v(a) \in D$.

Case II. *There is only one designated value, say* $D = \{0\}, N = \{1, 2\}$.

In this case we have (i) $\neg 0 = n$ (i.e. 1 or 2) by **L7**; $\neg 1 = 0$ or 2, $\neg 2 = 0$ or 1 by **L6**. If $\neg 1 = 2$ or $\neg 2 = 1$, then, by **L5** and **L4**, $1 \vee 2 = 2 \vee 1 = 0$. But then $\neg 0 \vee 1 = 0$ if $\neg 0 = 2$, and $\neg 0 \vee 2 = 0$ if $\neg 0 = 1$, contradicting **L1** in either case. Hence, (ii) $\neg 1 = 0, \neg 2 = 0$. By **L4**, (iii) $0 \vee x = x \vee 0 = 0$. By **L2**, (iv) $1 \vee 1 = n, 2 \vee 2 = n$. By (i) and **L1**, (v) $1 \vee 2 = 2 \vee 1 = n$.

At this stage, the separation of D and N is complete. The only remaining truth tables are of the form:

a	000	111	222
b	012	012	012
$a \vee b$	000	$0nn$	$0nn$
$\neg a$	n	0	0

But none of these tables can give **A4** a non-designated value. Thus, if $v(c)$ or $v(b)$ is 0, then **A4** gets 0. Therefore, $v(b), v(c) \in N$. If now $v(a) \in N$, then $v(\neg(c \vee a)) = 0$ and **A4** gets 0. If $v(a) = 0$, then $v(\neg(\neg a \vee b)) = 0$ and **A4** gets 0.

3.5 Recursive Functions

3.5.1 Recursive arithmetic

About 16 papers on recursive functions are included in this collection of Skolem's work. Of these, [1947a] deals with the history of the subject; the papers [1923a; 1939b; 1940a; 1950a; 1956a] in one way or another involve derivability and definability in formal systems. The rest deal with recursive functions as ordinary mathematical objects; they will be considered in the next section.

Skolem [1923a] "was written in the autumn of 1919, after I had studied the work of Russell and Whitehead". The purpose is to show that the use of free variables suffices to provide a foundation for large parts of mathematics. The justification for introducing bound " variables ranging infinite domains therefore seems very problematic; i. e., one can doubt the justification of the actual infinite or the transfinite". In [1947a], Skolem asserts that the paper [1923a] "is so far as I know the first investigation of recursive number theory. The utterance ... that recursive arithmetic can be traced back to Dedekind and Peano seems to me rather strange, because the researches of these men had another purpose than to avoid the use of quantifiers". Perhaps the development in Hermann Grassmann's *Lehrbuch der Arithmetik* [1861] might claim to be an (unconscious) partial anticipation of Skolem's paper; although the spirit of Skolem [1923a] is closer to Kronecker.

The 8 sections of [1923a] are: addition, less than, multiplication, divisible, subtraction and division, g. c. d. and l. c. m. , prime, and some explicit uses of bounded quantifiers. There is a polemic against regarding descriptions as incomplete symbols. "This conception does not seem to me to be beyond doubt; but even if it were correct for the descriptive functions of ordinary language, we would not need to adopt it for the descriptive functions of arithmetic". The culminating point

of the paper is a statement and proof of the unique prime factorization theorem (Theorem 74).

While Skolem's basis correspond roughly to (primitive) recursive arithmetic in Hilbert-Bernays [1934], Skolem does not give a formal system and in fact freely uses devices which can be seen directly to be justifiable by taking "the recursive mode of thought as basis".

Examples are

7.1 *Relations (propositional functions) can be defined recursively in the same manner as functions:*

Definition 2. $a < 1$ is false. $(a < b + 1) \equiv (a < b \vee a = b)$
Also definitions 5, 8, 12, 14, 15, 16, 18, 20.

7.2 *Bounded quantifiers are shown to be dispensable in several cases, e.g.*

Definition 5.

$$\begin{aligned} D(a,b,1) &\equiv (a = b) \\ D(a,b,c+1) &\equiv (D(a,b,c) \vee (a = b(c+1))) \\ b|a &\equiv D(a,b,a) \end{aligned}$$

But, in general, it is argued in §8 that we can freely use bounded quantifiers since we can see directly that they are all right. It is in Gödel [1931] that we have for the first time an explicit definition of (primitive) recursive functions and relations, as well as a general procedure for eliminating bounded quantifiers.

7.3 *Descriptions (with a restricted domain of existence) are treated informally in several definitions in a somewhat extended sense.*

Definition 7. Subtraction $(c - b = a) \equiv (c = a + b)$
Definition 9. Division $(c/b = a) \equiv (c = ba)$

7.4 *Definition by cases (using also course-of-values recursion implicitly).*

Definition 10. $a = b \supset (a, b) = a$; $a > b \supset (a, b) = ((a - b), b)$; $a < b \supset (a, b) = (a, (b - a))$.

7.5 *Double recursion of propositional functions.* Definitions 17 and 25.

7.6 *Definition of functionals.* Definition 23.

7.7 A bound function variable is used in the statement and proof of Theorem 76 to deal with the notion of cardinality. It is argued that the apparent vicious circle from a finitist point of view is actually harmless. "Thus even in the cases mentioned

we can first establish a theorem to the effect that the number of mappings that are possible at all is finite and after that give the general definition of the notion of cardinality for classes of objects".

Work since Skolem [1923a] by Ackermann, Gödel, Péter and others has rendered more explicit the relation between these diverse devices and the sharper domain of (primitive) recursive arithmetic. But we shall not pursue this matter.

From a philosophical point of view, Skolem [1923a] comes close to Hilbert's finitist position and is somewhat more restrictive than Brouwer's intuitionist viewpoint. In one of his papers in Spanish, Skolem asserts that his treatment of functions (10.7 above) is similar to Brouwer's free choice sequences.

Gödel's epoch-making paper of [1931] includes incidentally some central results on recursive functions. One of them is that all recursive functions are arithmetical in the sense of being definable in the ordinary system P of number theory which begins only with addition and multiplication as initial functions. (For a complete proof of this result, see Hilbert-Bernays [1934]). Skolem [1937a] gives an extension of this result to a wider class of inductively defined functions.

7.8 *If* $U(x_1, \ldots, x_n, y, z_1, \ldots, z_m)$ *is built by disjunction and conjunction out of given arithmetical functions of* $x_1, \ldots, z_m$ *and components of the form* $F(\xi_1, \ldots, \xi_n, \eta)$ *where* $\xi_1, \ldots, \xi_n, \eta$ *are such functions* $f_1, \ldots, f_n, f$ *of* $x_1, \ldots, z_m$ *that* $(\xi_1, \ldots, \xi_n)$ *always precedes* $(x_1, \ldots, x_n)$ *in the lexicographical order and the equations*

$$\xi_1 = f(x_1, \ldots, z_m), \ldots, \xi_n = f_n(x_1, \ldots, z_m), \eta = f(x_1, \ldots, z_m)$$

express arithmetical relations, and if

$$F(x_1, \ldots, x_n, y) \equiv \exists z_1 \ldots \exists z_m U(x_1, \ldots, x_n, y, z_1, \ldots, z_m)$$

then $F(x_1, \ldots, x_n, y)$ *is an arithmetical propositional function.*

The proof uses a device by which quantification over finite classes and relations are replaced by ones over natural numbers. The theorem is no longer true if negation is permitted in U. In fact, Skolem has previously given an example of the more general form which is not arithmetical. Alternatively, we can take from Hilbert-Bernays [1939] the definition of truth set of ordinary number theory P which, for obvious reasons, is not definable in P.

Skolem [1939b] shows that more general forms of free variable induction, such as

$$A(b, a) \text{ if } A(b, 0) \text{ and } A(\phi(b, a), a) \supset A(b, a + 1)$$

can be reduced to the ordinary form without using bound variables. An improvement of this is given in Péter [1940] (JSL, vol. 5, p. 35).

Skolem [1950a] contains the remark that in recursive arithmetic, at a certain stage, the usual rule of induction can be reduced to the special rule:

$$\text{If } f(0) = 0 \text{ and } f(n) \geq n + 1 \text{ , then } f(n) = 0$$

Skolem then asks whether it is even reducible to:

$$\text{If } f(0) = 0 \text{ and } f(n) = f(n') \text{ , then } f(n) = f(0)$$

The answer is yes and follows from a previous argument of Bernays [1945] (Proc. London Math. Soc. vol. 48, pp. 401-434).

Skolem [1956a] gives details of an alternative proof of a related result of Bernays (ibid., pp. 407-408) that in recursive arithmetic the induction rule follows from the principle that every function defined by the primitive recursive schema is uniquely determined.

3.5.2 Recursive function theory

The four notes of 1944 [1944a; b; c; d] and the paper [1953b] contain a number of observations concerning primitive and general recursiveness. The other four papers [1954b; 1954a; 1962b; 1963c] deal mainly with the problem of bases for recursively enumerable sets (a collection K of functions is said to form an r.e. basis if every r.e. set can be enumerated by a function in K).

In Skolem [1944a], the question is raised whether every general recursive function can be expressed in the simple form

(18) $\epsilon_y R(x_1, \ldots, x_n, y)$

where R is primitive recursive and $\forall x_1 \ldots \forall x_n \exists y R(x_1, \ldots, x_n, y)$.

8.1 *If ψ is monotone increasing in the familiar normal form $\psi(\epsilon_y S(x_1, \ldots, x_n, y))$ of f, then f can be expressed in the form (18).*

8.2 *Every general recursive function is the difference of two general recursive functions both expressible in the form (18).*

8.3 *A necessary and sufficient condition for a general recursive function f to be expressible in the form (18) is that $y = f(x_1, \ldots, x_n)$ is a primitive recursive relation. (Skolem [1944b]).*

8.4 *For suitable ϕ, ϕ is not primitive recursive, but $z = \phi(x, y)$ is.*

8.5 *It is not true that every general recursive functions with values restricted to 0 and 1 is a primitive recursive function.*

In the remainder of [1944b] and in [1944c], alternative proofs are given for the well-known results that there is no general method for deciding whether $\forall x \exists y P(x, y)$ for every primitive recursive P, and that there exists some primitive recursive P such that $\exists y P(x, y)$ is not primitive recursive.

Skolem [1944d] reverts to Theorem 9 in section 6 above from which it follows immediately that, for every sequence of number functions $f_1(x), f_2(x), \ldots$ there exist two functions $g(x)$ and $h(x)$, $g(x)$ monotone increasing, such that for every k, either $f_k(g(x)) > 0$ for all $x \geq h(k)$ or $f_k(g(x)) = 0$ for all $x \geq h(k)$. Results are then proved about these functions g and h:

8.6 *If the sequence $f_k(x)$ is an enumeration of all general (primitive) recursive functions of x, then $g(x)$ cannot be general (primitive) recursive.*

Skolem also asserts that if $f_k(x)$ is a recursive enumeration of all primitive recursive functions of x, then $g(h(x))$ is not recursive. The proof appeals to Gödel's incompleteness theorem but seems to require some additional assumptions of provability.

Skolem [1953b] contains six notes. The first note gives an alternative construction of an r.e. set which is not recursive. It is debatable whether this is simpler than familiar proofs. In note 2, it is remarked for each general recursive relation R, there are primitive recursive functions $f_1, \ldots, f_n$ such that

$$R(x_1, \ldots, x_n) = \exists t(x_1 = f_1(t) \vee \ldots \vee x_n = f_n(t))$$

In note 3, it is remarked that for those general recursive functions ϕ such that $y = \phi(x_1, \ldots, x_n)$ is primitive recursive, it is possible to replace them in quantifier-free contexts by primitive recursive functions, without introducing quantifiers. Note 4 mentions the fact concerning Gödel's incompleteness theorem that it is possible the $f(a) = g(a)$ may be true and yet $f(a) = 0$ is provable in a given system while $g(a) = 0$ is not.

In note 5, Skolem gives a short alternative proof of Markov's theorem that if $g(t)$ is a primitive recursive function which takes on all values infinitely often, then for each n, there is a primitive recursive predicate Q_n such that for each general recursive function f of $x_1, \ldots, x_n$ a number e can be selected to satisfy:

$$f(x_1, \ldots, x_n) = g(\mu_t Q_n(x_1, \ldots, x_n, t, e))$$

In note 6, it is shown that for every general recursive function $g(x)$, we can find some primitive recursive functions f and h, f assuming all values, such that $g(x) = h(\mu_t(x = f(t)))$. This result is actually due to Julia Robinson [1950] (Proc. Am. Math. Soc., vol. 1, pp. 213-221, §5).

Skolem [1954b] provides alternative formulation of Kalmar's class E of elementary functions and establishes that a number of basic devices can be obtained on these bases. In [1954a], Skolem suggests a narrower class H obtained from E by dropping the operation of indefinite products. It is shown that functions in H are

bounded by polynomials so that x^y is not in H. On the other hand $z = x^y$ can be expressed as an equation between functions in H. It is also shown that $m = p_n$ (p_n the n-th prime number) can be so expressed and that $(m)_n$ (the exponent of the highest power of p_n dividing m) is in H.

In [1962b] and [1963b], Skolem renames H the class of lower elementary functions and establishes:

8.7 *The lower elementary functions form an r.e. basis.*

Actually this follows from a result of Moh Shaw-Kwei [1956] (Acta Mathematica Sinica, vol. 6, pp. 548–564). But the proofs of Moh and Skolem are very different. At the end of [1962b], Skolem remarks: "It has been asked whether the recursively enumerable sets are all of them Diophantine sets ... I regret not having had the opportunity yet to study this question seriously".

4 SUPPLEMENTARY NOTES

4.1 The Algebraic tradition in the development of logic

It is usual to recognize two lines of development in the history of modern logic. Both take their point of departure the ideas of Leibniz of a *universal characteristic.* We cannot do justice to Leibniz's ideas in a short supplementary note. Let us just note that Leibniz's vision is seen by some as a principal source for the development of a formal or symbolic logic, a *calculus ratiocinator,* to systematize valid patterns of inference. For others, the real significance of the *universal characteristic* goes beyond formal logic and is seen as a *lingua philosophica* to express a knowledge of the real world.

Frege in his Begriffschrift explicitly declares his adherence to the *lingua philosophica* ideal and contrast his ideas with the Boole–Peirce–Schröder algebraic tradition, characterizing the latter as a mere *calculus raticinator;* see van Heijenoort [1967, p. 2].

I do not know how closely the later development followed the philosophical ideas of Leibniz. But there were two distinct schools in the development of modern logic. One the Frege-Peano-Russell-Hilbert tradition, the other the Boole-Peirce-Schröder tradition.

With the publication in 1928 of the first edition of the influential text-book of Hilbert–Ackermann, *Grundzüge der teoretischen Logik*, the Frege tradition came to dominate. And with the canonical source book, From Frege to Gödel [van Heijenoort, 1967], the algebraic approach to large extent disappeared from view. The heritage from Frege dominated the discussion, in particular, in issues of semantics and the theory of meaning; see e.g. the introductory chapter of Church [1956].

The history of logic is now becoming more faithful to the sources, and we see today an increased interest in understanding both lines of development and how

they interacted in the development of logic in the early part of the twentieth century.

A recent and interesting review of the development of the algebraic approach is the book by G. Brady, *From Peirce to Skolem, A Neglected Chapter in the History of Logic*, published in 2000. An important study of the landmark 1915 paper by L. Löwenheim, *Über Möglichkeiten in Relativkalkül*, is the book by C. Badisa, *The Birth of Model Theory*, from 2004. And a rich overview of the logic of Peirce can be found in the collection, *Studies in the Logic of Charles Sanders Peirce*, from 1997.

Löwenheim and Skolem emerge as central figures in these studies, integrating in their work the insights of both schools. But despite the great originality and importance of his 1915 paper, Löwenheim remained true to the algebraic tradition of Peirce and Schröder; see his late paper in the Journal of Symbolic Logic from 1940, *Einkleidung der Mathematik in Schröderschen Relativkalkül*.

Skolem became the genuine bridge-builder. His thesis for the state examination at the University of Oslo in 1913, *Undersøkelser indenfor logikens algebra*, was directly derived from Schröder volume I. The thesis was a rich and mathematically sophisticated study of the theory of classes, or in the language of Schröder, *das identische Kalkül*; see Skolem [1913b; 1919a; 1931d; 1936g].

Skolem spent the winter 1915–16 in Göttingen and became thoroughly familiar with the Hilbert school. At that time we know that the *Principia Mathematica* of Whitehead-Russell was intensively studied in Göttingen. This obviously had an influence on Skolem's work, as can be seen from his papers in the early 1920s; see, in particular, his remark in the paper Skolem [1923a], which was written in 1919 "after I had studied the work of Russell and Whitehead [Skolem 1923a, p. 187]".

After his great papers of the 1920s and early 1930s Skolem was recognized as a major figure in the development of mathematical logic. His early adherence to the Schröder tradition became less pronounced, e.g. in his lectures on logic at the University of Oslo in the mid-thirties (see [Skolem, 1936a]) he takes Hilbert and Ackermann [1928] as his model. There is no explicit mention of Schröder; but his early interests are present through a chapter on "the class calculus". In his later years he used to refer to the two volumes of Hilbert and Bernays as "the" text on mathematical logic.

4.2 Skolem and Lattice Theory

In the same year as his 1936 lectures we also find his paper, *Über gewisse "Verbände" or "lattices"*, see [Skolem, 1936g]. He takes note of the contemporary advances in this part of mathematics, in particular the work of G. Birkhoff, and he calls attention to his earlier work in the algebra of logic, Skolem [1913b; 1919a], where some of the "recent" result can already be found.

This paper was reviewed by Birkhoff [1937] (Journal of Symbolic Logic, vol 2). One gets the impression that the main point of the review is to to assign a marginal role to Skolem in the development of lattice theory. Wang in his survey of this

part of Skolem's work is rather non-committal on questions of priority.

Today we have a better picture. Workers in lattice theory have rediscovered some quite important ideas and results in the early papers of Skolem, including also in the part of Skolem [1920a] that was not reprinted in van Heijenoort [1967], nor mentioned by Skolem in [1936]. A brief account of this, with full references to the current literature, can be found in J. von Plato [2003], see also J. von Plato [2007]. Particularly noteworthy is Skolem's solution in [1920a] of the word problem for both free and finitely presented lattices through a polynomial time algorithm. This predates the work by P. Whitman [1941] on free lattices; for the rediscover see Burris [1995]. We also find in these early papers an anticipation of what later has become know as the Gödel-Dummett logic.

4.3 First Order Logic

The founders of modern logic, Frege, Schröder, Russell, and Zermelo, all worked within systems of higher-order logic. The story of how logic became predominantly a study of first-order systems is complex. It seems that it was Hilbert who in his lectures in 1917–18 was the first to make an explicit distinction between first order (*engere*) and higher order (*erweiterte*) systems of logic.

It was the work of Skolem and Gödel that gave prominence to first order logic. And since set theory can be axiomatised as a first order system, it was suggested that first order logic was sufficient as a universal frame-work. I use deliberately the words "it" and "frame-work" to hint at the complexity of the actual development; for a careful discussion see [Goldfarb, 1979] and the recent paper [Eklund, 1996]; the latter has a review of the literature and an extensive bibliography.

What is of interest for this discussion is the fact that Skolem never restricted himself to first-order logic. In his Oslo 1936 lectures he follows the tradition of Hilbert-Ackermann and has separate chapters on the *engere* and *erweiterte* functional calculus. And as late as in his lecture to the congress of the International Mathematical Union at Harvard University in 1950 he speaks about his recent studies on how to develop arithmetic in ramified type theory. He returns to the topic in his lectures on abstract set theory given at Notre Dame in 1957–56 (published in [Skolem, 1962c]). Skolem obviously took this system very seriously as a possible logical framework for arithmetic, and was "somewhat astonished by the general lack of interest in this theory [Skolem, 1950d]".

4.3.1 Löwenheim–Skolem Theorem

Löwenheim's theorem from 1915 and the later studies by Skolem in the 1920s are today recognized as the starting point for the model theory of first-order logic. Supplementing the discussion in Wang we draw attention to three later studies. The role of Skolem as a founder of model theory is convincingly argued in the survey article, *Model Theory before 1945*, by R. L. Vaught [1974]. A recent and

most important study is the book by C. Badesa [2004], *The Birth of Model Theory*. We should also mention again the book by G. Brady [2000], *From Peirce to Skolem; A Neglected Chapter in the History of Logic*. These works, and in particular, Badesa [2004], contain detailed studies of both the development of algebraic logic, the [1915] paper of Löwenheim, and the developments to come. The originality of Löwenheim's 1915 paper is striking, there is nothing in the previous development of logic that points to a result of this kind. One could speculate if he drew some inspiration from contemporary developments in mathematics. There were developments in algebra that could be suggestive, in particular W. Steinitz's [1910] great paper, *Algebraische Theorie der Körper*. I once suggested that the construction of the algebraic closure could have been a guiding example, and I noted that both Steinitz and Löwenheim were active in Berlin at about the same time. But, as reported in Brady [2000], p. 194, there is no evidence of any direct contacts between the two.

4.3.2 Completeness of First-order Logic

This is a topic that has been extensively discussed in the literature; a recent account can be found in the *Introductory Note* to Gödel's work on the completeness of first order logic by B. Dreben and J. van Heijenoort in Gödel [1986]. We also draw attention to Goldfarb's 1971 review of [Skolem, 1970], the Skolem lecture by Wang 1996, and the review of Gödel's correspondence as contained in volumes IV and V of his Collected Works by J. von Plato [2004] (Bulletin of Symbolic Logic, vol. 10, pp. 558–563).

We cannot go into details here, just point out that the fact that Skolem did not prove completeness of first-order logic is not necessarily an oversight on his part. It is, perhaps, more correct to say that Skolem and Gödel pursued different programs.

Skolem as a mathematician was interested in structures and algorithms. He was, however, thoroughly familiar with the axiomatic methods and formal proof systems; see his work on lattice theory referred to in Section 4.2. And his various proofs of the Löwenheim theorem showed an equal familiarity with the model theory of first-order logic. In his 1936 Oslo lectures he followed closely Hilbert-Ackermann [1928] in his presentation of logic as an axiomatic system. But even in the 1936 lectures there is no mention of the completeness theorem and its solution by Gödel [1930].

Skolem was following another program. Both in his studies of arithmetic, Skolem [1923a], and in his 1928 lecture on logic to the Norwegian Mathematical Society, Skolem [1928a], he was interested in algorithms and proof procedures. For him the axiomatic method was a tool, not a foundation. This was a view that was reinforced by the so-called "Skolem paradox" and his work on non-standard models for arithmetic.

Thus it was not unreasonable to adopt a strategy of working from below, i.e. on effective proof procedures in logic and on primitive recursion in arithmetic, and

to see how far one could go in catching logical and arithmetical "truth". In the 1920s this was a perfectly sound strategy; the limitations came in the 1930s with the work of Gödel [1931] and Church [1936].

Skolem took an immediate interest in the work of Herbrand [1931]. He even made a personal visit to the Herbrand family after Herbrand's death in order to obtain a copy of his thesis. The logical works of Herbrand are translated in [Herbrand, 1968], and they are provided with extensive comments by B. Dreben, W. Goldfarb, and J. van Heijenoort. Today we see the close relationship between the work of the Skolem, Gödel and Herbrand. Quoting from the Dreben–van Heijenoort Introductory Note [Gödel, 1986, p. 55] we may sum up the their relationship as follows: "Let us consider the sequence: Frege-type formal systems, proof procedures, set-theoretic validity. Skolem connects (at least implicitly) the second and the third, Herbrand the first and the second, Gödel the first and the third."

4.3.3 Entscheidungsprobleme

We make three short remarks.

(i) *The classical decision problem.* Wang, who was an expert himself in this field, presents in his survey a very complete picture. There is one new noteworthy result we may mention, Goldfarb's negative solution of the Gödel case with identity, which surely is a result that would have appealed to Skolem. An up-to-date account of the field can be found in Börger, Grädel and Gurevich [1997].

(ii) *Thue systems.* Skolem did some early work on decidability of mathematical systems. But it is interesting to note that he never dealt with the decidability questions raised by A. Thue [1914]. He has a brief mention of Thue's work in his 1936 Oslo lectures, but not identifying the problem as one concerning the word problem for semi-groups. Later, in Skolem [1954a], after he has become aware of Post [1947], he briefly reports the result. Post had been made aware of the problem by Church, who had an encyclopedic knowledge of all fields of mathematical logic. This is no great surprise, since Thue's problem fitted nicely in with the work of the Princeton schools. Skolem as a student of, and later as a younger colleague of Thue, surely knew the 1914 paper when it was published, but he never worked on the problem.

(iii) *Hilbert's 10th problem.* This is another story of opportunities lost. Toward the end of his life Skolem took note of the problem: "It has been asked whether the recursively enumerable sets are all of them Diophantine sets? I regret not having had the opportunity to study this question seriously [Skolem, 1962b]." His source was the well-known textbook by M. Davis from 1958.

Twenty years earlier, in 1938 Skolem published his influential book, *Diophantische Gleichungen.* The introduction is wholly concerned with various decision

problems for Diophantine equations, with due reference to Gödel [1931] and Church [1936] — not standard references in number theoretic texts at that time. He also notes, using the reduction techniques in Gödel [1931] and Skolem [1937a], what we recognize as the standard arithmetical representation of recursively enumerable sets. It "only" remains to reduce a mixed prefix to an existential form. Davis [1958] presents some fairly obvious steps towards this goal, steps that Skolem could have taken. This does not mean that Skolem could have taken the final steps. The solution of Hilbert's 10th problem belongs to Y. Matiyasevich.

4.3.4 The Epsilon Theorem

Here I want briefly to draw attention to the Skolem lecture on the epsilon-substitution method by G. Mints [1996]. Mints notes in his lecture that Wang in his survey of Skolem's work remarks that "it is hard to evaluate the importance of the proofs and comments in this section [Skolem, 1929b, §5]". What Mints points out is that this work "can be seen as instances of the epsilon-substitutions and the proof of the epsilon-theorems. They provide speed-up estimates".

4.4 Skolem and Axiomatic Set Theory

We have in recent years seen a number of studies on the history of axiomatic set theory as it developed in the early part of the twentieth century. The literature is vast, let me mention a few contributions of particular relevance to the understanding of Skolem's role in this development: the papers by A. Kanamori in the Bulletin of the ASL, see Kanamori [1996; 2004], and the forthcoming scientific biography of E. Zermelo by H.-D. Ebbinghaus [2007]. For a broader survey of the development of logic and set theory in the first part of the last century see Mancosu, Zach, and Badesa [2005].

These studies have given us a better understanding of how basic ideas and techniques emerged and developed into the axiomatic theory as we know it today. The work of Skolem on axiomatic set theory has its rightful place in these studies; in this note we add a few brief remarks supplementing the discussion in Wang's survey.

But before we continue the discussion of foundational matters, let us note that Skolem had a high opinion of Cantor as a creative mathematician; see Skolem [1962c]. It was above all Cantor's theory of ordinals that was the object of Skolem's admiration. At one place he writes that "these ordinals Cantor has introduced by a creative process which is very characteristic of his way of thinking" [Skolem, 1962c, p. 9]. This is Skolem speaking as a mathematician interested in structures and algorithms, not foundational matters. Skolem kept his interest in understanding the structures of ordinals and well-orderings "from below"; for a late contribution see Skolem [1956c]. The problem posed by Skolem in this paper is still a subject of research, see van den Dries L. and H. Levitz, [1984].

4.4.1 Definite Eigenschaften

As Skolem pointed out in his 1922 lecture, if you want to play the axiomatic game with respect to the set-theoretic foundation of mathematics, you can restrict yourself to a first-order formalism with the membership relation as the only non-logical symbol. This is a remarkable result, with Peano as one of the main sources. It was not at all obvious that the great wealth of mathematical concepts and constructs can be coded into a first order system with only one binary non-logical constant. And it is still a point to debate what, in addition to formal provability, is captured by this reduction. As we know from Ebbinghaus's biography, Zermelo had strong views about this issue; we shall return to this point in the remarks on his dispute with Skolem.

But if you accept to play the game as Skolem did in his 1922 paper, you can immediately give a precise answer to the troubling question how to express the notion of *definite Eigenschaften* within the axiomatization. Skolem's answer was to identify *definite Eigenschaften* with properties expressed by a first order formula with the membership-relation as the only non-logical symbol. Hermann Weyl had already in [1910] expressed similar ideas. His ideas were later developed in his book *Das Kontinuum* [1918]. This book is a landmark in the development of the predicative approach to the foundation of mathematics; for a careful historical analysis of Weyl's contribution and a survey of further developments see Feferman [1998].

Weyl and Skolem pursued different goals, but the technical details are similar. Thus, if Skolem was acquainted with the work of Weyl from 1910 and 1918, there should be an acknowledgments in his critical remarks from 1922. And it does indeed seem that Skolem knew about Weyl [1910]. The review of this work in the *Jahrbuch über die Fortschritte der Mathematik* published in 1913 carries the signature Sk. Sk has been read as an abbreviation for Skolem and the review has entered the literature as a publication by Skolem, see Boos [1995] and Mancosu, Zach, and Badesa 2005, where we in the latter paper read that "although Skolem does not mention Weyl [1910], he was familiar with it, as he had reviewed it for the Jahrbuch". And in the bibliography we find an item: Skolem 1912. But this is not correct, a closer inspection of the *Jahrbuch* reveals the Sk is an acronym for professor Salkowski, a geometer from Berlin. Concerning the 1922 lecture, this is material originating from his visit to Göttingen in 1915–16 after Weyl had left for Zürich, and thus independent of Weyl [1918].

It is, however, noteworthy that Skolem in his later work on the constructive foundation of mathematics only has a few superficial references to Weyl. We will return to this point in Section 4.6.

4.4.2 The Skolem Paradox

The so-called Skolem Paradox has an extensive literature. Supplementing the discussion in Wang's survey, we mention some recent contributions: A careful analysis of the mathematical content of Skolem's argument can be found in the PhD thesis

by T. Bays [2000]; see his article in the *Handbook of the Philosophy of Science*, volume 5. For some recent studies of the philosophical and foundational implications of the paradox see A. George, *Skolem and the Löwenheim Theorem: A Case Study of the Philosophical Significance of Mathematical Results* [George, 1985], I. Jané [2001], *Reflections on Skolem's Relativity of Set-Theoretical Concepts*, and F.A.Muller [2004], *Deflating Skolem*.

For the working set theorist of today the Skolem Paradox has been turned into a standard technical tool, the Skolem Hull construction, to be listed among the preliminaries to further studies; see as an example, [Kanamori, 1997, p. 8].

4.4.3 The Dispute with Zermelo

Zermelo reacted strongly against the Skolem 1922 lecture, both to the Löwenheim–Skolem theorem and the application of it to axiomatic set theory. A full account of the so-called "Skolem Controversy" can be found in Ebbbinghaus [2007, sections 4.8 and 4.9]. Zermelo evidently saw a challenge to real mathematics in the "Skolem relativism", and he started a campaign against it — a war on *Skolemism* in the words of the German mathematician R. Baer (see [Ebbinghaus, 2007, 4.8.3]).

Except for one exchange, the Zermelo [1929] paper and the response in Skolem [1930b], Skolem did not participate in the war. In the 1930s he was occupied with his work on algebra, number theory, the decision problem, and the construction of nonstandard models for arithmetic. The latter work can, perhaps, implicitly be read as an argument against Zermelo. It is also possible to read Skolem's 1938 Zürich lecture in the same way, but only indirectly since there is no explicit reference to Zermelo beyond the classics of 1908.

Both Skolem and Zermelo were logicians of the first rank, and we need to pay attention to the deep insights of both men. And, indeed, we do, most logicians working in axiomatic set theory today are guided by the insight of Zermelo, both in his fundamental studies from the early years of the century and the later papers of the 1930s, in particular, his *Über Grenzzahlen und Mengenbereiche*. But when we want to turn insights into proofs, we use the first-order encoding and tools such as the Skolem Hull Construction, a latter day technical version of the Löwenheim-Skolem theorem.

4.5 Nonstandard Models of Arithmetic

Skolem did not believe in an axiomatic foundation for mathematics. He could and did play the axiomatic game on many occasions, and he never mistrusted the axiomatic method as a tool in the study of mathematical structures. But axioms can never uniquely characterize the concepts involved. This, in his opinion, was a direct consequence of the Löwenheim-Skolem theorem in the case of set theory; see the discussion in Section 4.4. He quite early suspected that the same "relativism" would also apply to the natural number sequence; see e.g. Skolem [1929b, p. 269], where he writes "It would be interesting to verify that every collection (Sammlung) of sentences (Sätze) expressed in the language of first-order

arithmetic (Zahlaussagen) remains valid even when one makes certain changes to the interpretation (Deutung) of "number". If my conjecture is correct, one can expect that it is always possible to make such reinterpretations without destroying the validity (ohne Zerstörung der Gültigkeit) of the sentences." During the 1920s and early 1930s he published a number of works towards this goal. And in [1933d] and [1934a] he succeeded in the construction of a nonstandard model of arithmetic.

Today the study of nonstandard models is an important part of logic and mathematics. And what was originally intended as a critique, has been turned into an efficient tool in logic and analysis. For a recent survey of applications to the metatheory of first order arithmetic, see Hájek and Pudlak [1993]. For applications to nonstandard analysis, see the founding memoir, A. Robinson 1966.

Nonstandard analysis is yet another case where Skolem could have taken an important step forward, but did not. Skolem argued in a lecture to a Scandinavian congress in 1929 against using infinitesimals as a foundation for analysis: "We know that such entities can be introduced — they can be exhibited in so-called non-archimedian number systems — but it is not possible to construct a calculus of infinitesimals on such a foundation." But history does not always follow preconceived ideas. In 1933–34 Skolem constructed a non-archimedian, elementary extension of the natural numbers. This is, in the same way as the Skolem Paradox, no contradiction, one only needs to make a distinction between internal and external properties of the model. Thirty years later A. Robinson observed that if one inverts an infinitely large number in the Skolem model, one has a genuine infinitesimal. And he then proceeded to create a powerful system of infinitesimal analysis, see Robinson [1966].

Today we construct nonstandard models using the ultraproduct method. Skolem used a more direct combinatorial approach. And he hoped that he could make his method constructive, see the concluding remarks in Skolem [1955d]. But this is not so, a careful comparison of Skolem's construction and the ultraproduct construction was carried out by D. Scott [1960]. A clear exposition is given in Bell and Slomson [1969]. The crucial combinatorial lemma (see theorem 9 in Wang's survey) is today proved by an ultraproduct construction, but the combinatorial methods of the two proofs are basically the same. And contrary to Skolem's hopes, the arguments go beyond the constructive (recursive) as they essentially depend upon the Weak König Lemma.

4.6 Foundational Matters

The words "Begründung", "Grundlagenfragen", and "foundational research" occur frequently in the titles of Skolem's papers. But to which extent did Skolem have a complete and coherent view of the foundation of mathematics?

The young Skolem was a mathematician in the tradition of Kronecker; see e.g. his discussion of Kronecker in Skolem [1955c]. The active professors of mathematics in Oslo in Skolem's student days were A. Thue and C. Størmer, both experts in number theory, in particular, in solving diophantine equations. Skolem continued

in this tradition, in fact, it is possible, as argued in Jervell [1996], to see this theme, solving equations, as a common concern in much of Skolem's work, including logic. His fellow student was Viggo Brun, who later became famous for his work in combinatorics (on the Goldbach conjecture). As later colleagues, Skolem and Brun edited a new edition of the standard text on combinatorics, Netto's *Lehrbuch der Combinatorik*.

The young Skolem was as mathematician interested in structures and algorithms, working on the algebra of logic and on combinatorics. He knew, of course, the new and exiting work of Zermelo on the axiom of choice and the foundation of set theory. But he was skeptical as can be seen from his critical attitude to the Zermelo axiomatization of set theory expressed during his Göttingen winter of 1915–16.

Skolem was himself an expert in using the axiomatic method in the study of algebraic structures, see his early work on lattice theory, Section 4.2. But he did not believe in the possibility of an axiomatic characterization of basic concepts such as *numbers* and *sets*. His skepticism was based on his work on the so-called "set theoretical relativism" and on his construction of nonstandard models for arithmetic. This was his firm conviction, and it is not likely that the current discussion of the Skolem Paradox would have changed his attitude towards foundational matters.

His first publication on recursive mathematics was the 1923 paper on the *Begründung der elementaren Arithmetik durch die rekurrierende Denkweise*. This was the starting step — and it remained so. Skolem never worked out a complete account similar to what we find in the work of R. Goldstein [1957] and [1961] on recursive arithmetic.

Skolem did, however, return to his "program" on many later occasions, some important references are Skolem [1950a] and [1955c], but then mostly in form of examples and suggestions. He wanted step-wise to build up a "sound" foundation. This is clear from the pioneering Skolem [1923a], and is explicit in his later lecture on *The logical background of arithmetic*, Skolem [1953a]. His style bears some resemblance to current work in computer science, and some have seen him as an early pioneer of this field, see Jervell [1996].

"Basic objects" were almost always numbers. Skolem did not make any systematic attempt to build up a "constructive" analysis. He made over the years a number of references to Brouwer and intuitionism, but it is unclear how well he was acquainted with Brouwer's actual writings. In his lectures on set theory at Notre Dame, Skolem [1962c], he includes several short sections on alternative approaches such as intuitionism and his own work on ramified type theory. It is further of interest to note that he has a substantial section on the "operative" approach of P. Lorenzen. This was an approach that evidently appealed to Skolem.

Skolem was acquainted with the work of H. Weyl [1918] on predicative analysis. But he never seemed to have made a thorough study of Weyl's work, and there are only some very superficial references to it in his papers. This is, perhaps, due to a difference in style. Both Skolem and Weyl were concerned with the problem of impredicativity and the unrestricted use of quantifiers in the tradition of Cantor

and Frege. Skolem wanted to build bottom-up. Weyl, both in his [1910] paper and his [1918] book, continued to work in the axiomatic tradition, but with proper restrictions on the set existence axioms. In particular, he restricted the axiom of comprehension to arithmetical properties, i.e. to properties expressible using only first-order quantifiers. A comprehensive study of Weyl's contributions can be found in Feferman [1998], see, in particular, the chapter *Weyl vindicated: Das Kontinuum seventy years later*.

Skolem was, as we have argued, not interested in the meta-mathematics of the constructive approach. It is interesting to note a similar attitude on Brouwer's part. He did not e.g. regard the axiomatization of intuitionistic logic to be particularly important, what mattered was the actual working out of the mathematics guided by the correct insights.

Today we have a rather complete view of the meta-mathematics of the many constructive approaches to mathematics, an excellent reference is Feferman [1998]. How would Skolem fit into the picture drawn by Feferman? Skolem worked with primitive recursive arithmetic in his [1923a] paper, he used some form of the König lemma in his construction of non-standard models of arithmetic, he studied various extended schemes of recursion. But as he did not develop his ideas into a full account, we cannot know what principles he ultimately would have accepted. But he would have been pleased to see the extent of "real" mathematics that can be developed on a recursive or predicative basis.

5 SCIENTIFIC BIBLIOGRAPHY OF THORALF SKOLEM

Items in the bibliography are indexed by giving the year in which they appeared, and for each year the various titles are distinguished by letters a, b, and so on. As far as possible, an attempt has been made to follow the actual chronological order. For lectures, the year given is usually that in which they were delivered.

A * in front of an entry indicates that this item is included in [Skolem, 1970].

Each entry includes references to reviews of it in the current reviewing journals. Here have also been given references to "From Frege to Gödel. A source book in mathematical logic" (edited by J. van Heijenoort, Harvard University Press), for papers reprinted there since each paper in this collection has been provided with an extensive introductory comment.

The abbreviations used are mostly standard. Some of the lesser known are:

K.V. Skr.	for Skrifter, Videnskabsakademiet i Kristiania.
V.A. Skr.	for Skrifter, Vitenskapsakademiet i Oslo.
V.A. Avh.	for Avhandlinger, Vitenskapsakademiet i Oslo.
K.N.V. Forh.	for Kongelige Norske Videnskabsselskabs Forhandlinge, Trondheim.
N.M.F. Skr.	for Norsk Matematisk Forening, Skrifter.
N.M.T.	for Norsk Matematisk Tidsskrift.
C.M. Inst.	for Christian Michelsens Institutt, Bergen.

The reviewing journals are identified by the following abbreviations:

MR for Mathematical reviews.
Z for Zentralblatt für Mathematik.
R for Referativnyĭ Žurnal.
JSL for Journal of Symbolic Logic.
J for Jahrbuch der Mathematik.

1913

*b. Om konstitusjonen av den identiske kalkyls grupper.
(On the structure of groups in the identity calculus.)
Proceedings of the 3rd Scand. Math. Congress. Kristiania 1913, pp 149-163.

1919

*a. Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen.
K.V. Skr No 3, pp 37.

1920

*a. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit und Beweisbarkeit mathematischen Sätze nebst einem Theoreme über dichte Mengen.
K.V. Skr. I, No 4, pp 1-36.
J 48, p 1121 (Fr. Lange-Nielsen).
From Frege to Gödel pp 252-263.

1922

*b. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre.
Proceedings of the 5th Scand. Math. Congress. Helsinki 1922, pp 217-232.
J 49, p 138 (Fraenkel).
From Frege to Gödel pp 290-301.

1923

*a. Begründung der elementären Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich.
K.V. Skr. I, No 6, 38 pp.
From Frege to Gödel pp 302-333.

1924

b. Ein Verfahren zu beliebig angenäherter Bestimmung einer Wurzel einer beliebigen algebraischen Gleichung.
N.M.F. Skr. Series 1, No 15, 14 pp.

1926

a. Litt om de viktigste diskussioner i den senere tid angående matematikkens grundlag. (The most important recent discussions on the foundation of mathematics.)
N.M.T. 8, pp 1-13.
J 52, p 51 (Mollerup).

1927

*c. *Netto: Lehrbuch der Combinatorik.*
2nd ed. by Viggo Brun and T. Skolem.
Leipzig 1927.
(The 2nd edition contains a new chapter and several notes, totalling 60 pages, written by Skolem.)
J 53, p 73 (G. Feigl).

1928

*a. Über die mathematische Logik.
N.M.T. 10, pp 125-142.
J 54, p 58 (Feigl).
From Frege to Gödel, pp 508-524.

1929

*a. Über die Grundlagendiskussionen in der Mathematik.
Proceedings of the 7th Scand. Math. Congress, Oslo 1929, pp 3-21.
J 56, p 822 (Feigl).

*b. Über einige Grundlagenfragen der Mathematik.
V.A. Skr. I, No 4, pp 1-49.
J 55, p 31 (Fraenkl).

1930

*b. Einige Bemerkungen zu der Abhandlung von E. Zermelo: "Über die Definitheit in der Axiomatik".
Fund. Math. T XV, pp 337-341.
J 56, p 46 (Scholz).

*c. Über einige Satzfunktionen in der Arithmetik.
V.A. Skr. I, No 7, pp 1-28.
J 57, p 1320 (Ackermann).
Z 2, p 3 (Gödel).

1931

a. Den matematiske logikk og matematikken.
(The mathematical logic and the arithmetic.)
C.M. Inst. I, 2, 20 pp.
J 57, p 1317 (Skolem).

*d. Über die symmetrisch allgemeinen Lösungen im identischen Kalkul.
V.A. Skr. I, No6, pp 1-32.
J 58, p 65 (Ackerman).
Z 4, p 385 (Gödel).

1933

*c. Ein kombinatorischer Satz mit Anwendung auf ein logisches Entscheidungsproblem.
Fund. Math. T XX, pp 254-261.
J 59, p 54 (Ackermann):
Z 7, p 97 (Gödel).

*d. Über die Unmöglichkeit einer Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems.
N.M.F. Skr. Series 2, No 1-12, pp 73-82 and C.M. Inst. No 24.
J 59, p 53 (Ackermann).
Z 7, p 193 (Gödel).

1934

*b. Über die Nichtcharakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariabeln.
Fund. Math. T XXIII, pp 150-161.
J 60, p 25 (Ackermann).
R 1963, 12A70 (E. Orlovskiĭ).
Z 10, p 49 (Gödel).

d. Den matematiske grundlagsforskning.
(Foundational research in mathematics.)
N.M.T. 16, pp 75-92.
J 60, p 842 (Pietsch).

1935

*a. Ein Satz über Zählausdrücke.
Acta Szeged. T7, pp 193-199.
J 61, p 973 (Kalmár).
Z 12, p 385 (H.B. Curry).

*f. Ein Satz über die Erfüllbarkeit von einen Zahlausdrücke der Form $(x)(Ey_1, \ldots, y_n)K_1(x, y_1, \ldots, y_n) \,\&\, (x_1, x_2, x_3)K_2(x_1, x_2, x_3)$.
V.A. Skr. I, No 8, pp 1-10.
J 62, p 42 (Ackermann).
Z 13, p 243 (H.B. Curry).

*g. Über die Erfüllbarkeit gewisser Zählausdrücke.
V.A. Skr. I, No 6, pp 1-14.
J 61, p 973 (Kalmár).
JSL 1, p 111 (C.H. Langford).
Z 13, p 97 (H.B. Curry).

1936

a. Utvalgte kapitler av den matematiske logikk.
(Selected chapters of mathematical logic.)
C.M. Inst. VI, 6, 73 p.
J 62, p 1048 (Ackermann).
JSL 2, p 87 (Jørgen Jørgensen).

e. Eine Bemerkung zum Entscheidungsproblem.
Proceedings of 8th International Math. Congress, Oslo 1936.
J 63, p 31.
JSL 3, p 57.

*f. Einige Reduktionen des Entscheidungsproblems.
V.A. Skr. I, No 6, pp 1-17.
J 62, p 1059 (Kalmár).
JSL 2, p 86 (C.H. Langford).
Z 15, p 338 (H.B. Curry).

*g. Über gewisse “Verbände” oder “Lattices”.
V.A. Skr. I, No 7, pp 1-16.
J 62, p 1099 (Kalmár).
JSL 2, p 50 (Birkhoff).
Z 15, p 243 (Birkhoff).

1937

*a. Über die Zurückführbarkeit einiger durch Rekursionen definierter Relationen auf "Arithmetische".
Acta Szeged. T8, pp 73-88.
J 63, p 25 (Ackermann).
JSL 2, p 85 (Barkley Rosser).
Z 16, p 194 (H.B. Curry).

1938

c. *Diophantische Gleichungen.*
Zentralblatt für Mathematik B5, 1938, 130 p.
and Chelsea Publishing Company, Bronx.
J 64, p 112 (Boehle).
Z 18, p 293 (Mahler).

1939

*b. Eine Bemerkung über die Induktionsschemata in der rekursiven Zahlentheorie.
Monatsheft für Math. und Physik, Leipzig und Wien, 48, pp 268-276.
J 65, p 31 (Ackermann).
Z 21, p 385 (Hermes).
MR 1, p 132 (Frink).

1940

*a. Einfacher Beweis der Unmöglichkeit eines allgemeinen Lösungsverfahren für Arithmetische Probleme.
K.N.V. Forh. XIII, pp 1-4.
J 66, p 32 (Ackermann).
JSL 9, p 21 (Rózsa Péter).
MR 2, p 210 (Frink).

1941

*a. Sur la Porté du Théorème de Löwenheim-Skolem.
Les Entretiens de Zürich, 6-9 Dec. 1938, Zürich 1941, pp 25-47, discussion pp 47-52.
MR 2, p 338 (Curry).
JSL 7, p 35 (Saunders MacLane).

1944

*a. Remarks on recursive functions and relations.
K.N.V. Forh. XVII, No 22, pp 89-92.
JSL 11, p 26 (Paul Bernays).
MR 8, p 4 (Curry).

*b. Some remarks on recursive arithmetic.
K.N.V. Forh. XVII, No 26, pp 103-106.
JSL 11, p 26 (Paul Bernays).
MR 8, p 4 (Curry).

*c. A note on recursive arithmetic.
K.N.V. Forh. XVII, No 27, pp 107-109.
JSL 11, p 26 (Paul Bernays).
MR 8, p 4 (Curry).

*d. Some remarks on the comparison between recursive functions.
K.N.V. Forh. XVII, No 32, pp 126-129.
JSL 11, p 26 (Paul Bernays).
MR 8, p 4 (Curry).

1946

a. Den rekursive aritmetikk.
(Recursive arithmetic.)
N.M.T. 28, pp 1-12.
JSL 13, p 169 (Oiva Kentonen).
MR 8, p 125 (Curry).

1947

*a. The development of recursive arithmetic.
Dixième Congres des Math. Scand. Copenhagen 1946, pp 1-16, Copenhague 1947.
JSL 13, p 54 (David Nelson).
MR 8, p 558 (Curry).
Z 30, p101 (Ackermann).

1950

*a. A remark on the induction scheme.
K.N.V. Forh. XXIII, No 36, pp 167-170.
JSL 16, p 220 (Paul Bernays).
MR 11, p 635 (Curry).
Z 36, p 7 (Lorenzen).

c. De logiske paradokser og botemidlene mot dem.
(The logical paradoxes and remedies against them.)
N.M.T. 32, pp 2-11.
JSL 16, p 62 (G.H. von Wright).
MR 11, p 708 (Curry).
Z 40, p 294 (G. Hasenjaeger).

*d. Some remarks on the foundation of set theory.
Proceedings of International Congress of Math., Masschussets 1950, pp 695-704.
JSL 18, p 77 (Abner Shimony).
MR 13, p 521 (R.M. Martin).
Z 49, p 165 (J.C. Shepherdson).

1951

*a. On the proof of independence of the axioms of the classical sentential calculus.
K.N.V. Forh. XXIV No 6, pp 20-25.
JSL 18, p 67 (Alonzo Church and Nicholas Rescher).
MR 14, p 526 (R.M. Martin).
Z 47, p 13 (A. Rose).

1952

*b. A remark on a set theory based on positive logic.
K.N.V. Forh. XXV, No 22, pp 112-116.
JSL 21, p 97 (Frederic B. Fitch).
MR 14, p 937 (Novak Gál).
Z 50, p 246 (J.C. Shepherdson).

1953

*a. The logical background of arithmetic.
Bull. de la Société Math. de Belgique 1953, pp 23-34.
JSL 23, p 228 (H.G. Rice).
MR 16, p 553 (G. Kreisel).
R 1956, 8528 (T.L. Maĭstrova).
Z 57, p 6 (D. Tamari).

*b. Some considerations concerning recursive functions.
Math. Scand. 1, pp 213-221.
JSL 21, p 98 (Davis).
MR 15, p 667 (Frink).
R 1956, 964 (B.A. Trahtenbrot).
Z 52, p 248 (J.C. Shepherdson).

d. Resultater i grundlagsforskningen.
(Results in foundational research.)
Proceedings of Math. Congress, Lund 1953, pp 273-289.
MR 16, p 553 (Curry).
R 1958, 3512 (A.S. Esenin-Vol'pin).
Z 56, p 254 (H. Gericke).

1954

*a. Some considerations concerning recursive arithmetic.
Bull. de la Société Math. de Belgique 1954, pp 35-46.
JSL 23, p 228 (H.G. Rice).
MR 16, p 553 (Kreisel).
Z 56, p 248 (J.C. Shepherdson).

*b. Remarks on "Elementary" arithmetic functions.
K.N.V. Forh. XXVII, No 6, 6 p.
MR 16, 324 (Novak-Gál).
R 1956, 963 (B.A. Trahtenbrot).
Z 58, p 5 (H. Ostmann).

e. A critical remark on foundational research.
Proceedings of International Congress of Math., Amsterdam 1954,
Short lectures, pp 408-409.
R 1955, 5572 (T.L. Maĭstrova).

1955

*c. A critical remark on foundational research.
K.N.V. Forh. XXVIII, No 20, pp 100-105.
MR 17, p 817 (P.C. Gilmore).
R 1958, 945 (V.A. Uspenskiĭ).
Z 66, p 248 (H. Scholz).

*d. Peano's axioms and models of arithmetic.
Studies in logic and the foundation of Math., Amsterdam 1955, pp 1-14.
JSL 22, p 306 (S. Feferman).
MR 17, p 699 (E. Mendelson).
R 1958, 59 (V. Detlovs).

e. The logical nature of arithmetic.
Synthese 9, pp 375-384.
JSL 23, p 237 (A. Church).
MR 18, p 269 (H.B. Curry).
R 1956, 1927 (B.A. Trahtenbrot).

1956

*a. A version of the proof of equivalence between complete induction and the uniqueness of primitive recursions.
K.N.V. Forh. XXIX, No 3, pp 10-15.
JSL 24, p 69 (R. Péter).
MR 18, p 2 (G. Kreisel).
R 1967, 52 (È.S. Orlovskiĭ).
Z 72, p 6 (J.C. Shepherdson).

c. An ordered set of arithmetic functions representing the least ε-number.
K.N.V. Forh. XXIX, No 12, pp 54-59.
MR 18, p 785 (S. Ginsburg)
R 1959, 6536 (A. S. Esenin-Vol'pin).

1957

*b. Two remarks on set theory.
Math. Scand. 5, pp 40-46.
MR 19, p 933 (E.J. Cogan).
R 1959, 1371 (A.S. Esenin-Vol'pin).
Z 79, p 76 (J.C. Shepherdson).

d. Mengenlehre gegründet auf einer Logik mit unendlich vielen Wahrscheinlichkeitswerten.
S-Ber. Berliner Math. Ges. 1957/58, pp 41-56.
MR 27, 3511 (J.R. Schoenfield).
R 1961, 1A84 (Medvedev).
Z 116, p 9 (A. Oberschelp).

*e. Bemerkungen zum Komprehensionsaxiom.
Zeitschrift für Math. Logik und Grundlagen der Math. 3, pp 1-17.
JSL 32, p 128 (A. Lévy).
MR 20, 3067 (H.B. Curry).
R 1959, 6545 (A.S. Esenin-Vol'pin).
Z 95, p 9 (W. Ackermann).

f. *Forelesninger i mengdelære.*
(Lectures on set theory.)
Universitetets mat. inst. Oslo 1957, 57 p (mimeographed).

1958

*a. Une relativisation des notions mathématiques fondamentales.
Colloques internationaux des Centre National de la Recherche Scientifique, Paris 1958, pp 13-18.
MR 21, 4895 (E.W. Beth).
Z 87, p 8 (E. Specker).

*b. Remarks on the connection between intuionistic logic and a certain class of lattices.
Math. Scand. 6, pp 231- 236.
MR 21, 5574 (D. Scott).
R 1961, 2A56 (F.A. Kabakov).
Z 92, p 5 (B. van Rootselaar).

*d. Reduction of axiom systems with axiom schemes to systems with only simple axioms.
Dialectica 12, pp 443-450.
JSL 27, p 230 (G. Hasenjaeger).
MR 21, 3325 (R.M. Martin).
R 1959, 9696 (S.I. Adjan).
Z 93, p 12 (G. Takeuti).

1960

*b. A set theory based on a certain 3-valued logic.
Math. Scand. 8, pp 127-136.
MR 23, A 3669 (A. Rose).
R 1961, 11A84 (F.A. Talaljan).
Z 96, p 243 (W. Ackermann).

*c. Investigations on a comprehension axiom without negation in the defining propositional functions.
Notre Dame Journal of formal logic 1, pp 13-22.
MR 26, 4890 (K. Schütte).
Z 116, p 10 (S. Maehara).

1961

*a. Interpretation of mathematical theories in the first order predicate calculus.
Festschrift vor A. Fraenkel 1961, pp 218-225.
MR 29, 4671 (H. Ribeiro).

1962

*a. Recursive enumeration of some classes of primitive recursive functions and a majorisation theorem.
K.N.V. Forh. XXV, No 26, pp 142-148.
MR 27, 3512 (P. Axt).
R 1964, 1A111 (V. Šurygin).
Z 106, p 239 (E. Burger).

*b. Proof of some theorems on recursively enumerable sets.
Notre Dame Journal of formal logic 3, pp 65-74.
MR 27, 3513a (P. Axt).
Z 113, p 6 (T. Hailperin).

c. *Abstract set theory.*
Notre Dame Math. Lectures No 8, v+70 p.
MR 28, 20 (P. Lorenzen).

1963

*b. Addendum to "Proof of some theorems on recursively enumerable sets".
Notre Dame Journal of formal logic 4, pp 44-47.
MR 27, 3513b (P. Axt).
R 1965, 11A81 (G. Minc).

*c. Studies on the axiom of comprehension.
Notre Dame Journal of formal logic 4, pp 162-170.
MR 29, 5716 (E. Mendelson).
Z 133, p 258 (Fenstad).

1970

Selected Works in Logic
Universitetsforlaget, Oslo, 732 p.

BIBLIOGRAPHY

[Badesa, 2004] C. Badesa. , The Birth of Model Theory, Princeton University Press, 2004.
[Bays, to appear] T. Bays. The mathematics of Skolem's paradox, in Handbook of the Philosophy of Science, vol. 5, pp 485-518, to appear.
[Bell and Slomson, 1969] J. L. Bell and A. B. Slomson. Models and Ultraproducts, North Holland Publ. Comp., Amsterdam, 1969.
[Birkhoff, 1937] G. Birkhoff. Review of [Skolem, 1936g[, Journal of Symbolic Logic, vol. 2, pp 50-51, 1937.
[Boos, 1995] W. Boos. Thoralf Skolem, Hermann Weyl and "Das Gefühl der Welt als begrenztes Ganzes", in J. Hintikka, ed, From Dedekind to Gödel, Klüwer Academic Press, Dordrecht, pp 283-329, 1995.
[Brady, 2000] G. Brady. From Peirce to Skolem: A Neglected Chapter in the History of Logic, Elsevier, Amsterdam, 2000.
[Burris, 1995] S. Burris. Polynomial Time Uniform Word Problems, Mathematical Logic Quarterly, vol. 41, pp 173-182, 1995.
[Börger *et al.*, 1997] E. Börger, E. Grädel and Y. Gurevich. The Classical Decision Problem, Springer, Heidelberg, 1997.
[Church, 1936] A. Church. A note on the Entscheidungsproblem, Journal of Symbolic Logic, vol. 1, pp 40-41 and 101-102, 1936.
[Church, 1956] A. Church. Introduction to Mathematical Logic, Princeton University Press, 1956.
[Davis, 1958] M. Davis. Computability and Unsolvability, McGraw-Hill, New York, 1958..
[Dreben and van Heijenoort, 1986] B. Dreben and J. van Heijenoort. Introductory note to Gödel 1929, 1930 and 1930a, in Gödel vol. I, pp 44-59, 1986.
[Ebbinghaus, 2007] H. D. Ebbinghaus, in cooperation with V. Peckhaus. *Zermelo. An Approach to his Life and Work*. Springer, Berlin, 2007.
[Eklund, 1996] M. Eklund. On How Logic Became First-Order, Nordic Journal of Philosophical logic, vol. 1, pp 147-167, 1996.
[Feferman, 1998] S. Feferman. In the Light of Logic, Oxford University Press, 1998.
[George, 1985] A. George. Skolem and the Löwenheim-Skolem Theorem: A Case Study of the Philosophical Significance of Mathematical Results, History and Philosophy of Logic, vol. 6, pp 75-89, 1985.
[Goldfarb, 1971] W. Goldfarb. Review of [Skolem, 1970], The Journal of Philosophy, vol. 68, pp 520-530, 1971.

[Goldfarb, 1979] W. Goldfarb. Logic in the Twenties: the Nature of the Quantifier, Journal of Symbolic Logic, vol. 44, pp 351-368, 1979.
[Gödel, 1931] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatsheft für Mathematik und Physik, vol. 38, pp 173-198, 1931.
[Gödel, 1986] K. Gödel. Collected Works, vol. I, S. Feferman et al, eds., Oxford University Press, 1986.
[Goodstein, 1957] R. L. Goodstein. Recursive Number Theory, North Holland Publ. Comp., Amsterdam, 1957.
[Goodstein, 1961] R. L. Goodstein. Recursive Analysis, North Holland Publ. Comp., Amsterdam, 1961.
[Hájek and Pudlak, 1993] P. Hájek and P. Pudlak. Meta-Mathematics of First-Order Arithmetic, Springer, Heidelberg, 1993.
[Herbrand, 1971] J. Herbrand. Logical Writings, ed. W. Goldfarb, D.Reidel Publ. Comp., Dordrecht, 1971.
[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. Grundzüge der theoretischen Logik, Springer, Heidelberg, 1928.
[Hilbert and Bernays, 1934/1939] D. Hilbert and P. Bernays. Grundlagen der Mathematik, Springer, Heidelberg, 1934/1939.
[Houser *et al.*, 1997] N. Houser, D. D. Roberts and J. van Evra, eds. Studies in the Logic of Charles Sanders Peirce, Indiana University Press, 1997.
[Jané, 2001] I. Jané. Reflections on Skolem's Relativity of Set-Theoretical Concepts, Philosophica Mathematica, vol. 9, pp 129-153, 2001.
[Jervell, 1996] H. R. Jervell. Thoralf Skolem: Pioneer of Computational Logic, Nordic Journal of Philosophical Logic, vol. 1, pp 107-117, 1996.
[Kanamori, 1996] A. Kanamori. The Mathematical Development of Set Theory from Cantor to Cohen, Bulletin of Symbolic Logic, vol. 2, pp 1-71, 1996
[Kanamori, 1997] A. Kanamori. The Higher Infinite, Springer, Heidelberg, 1997..
[Kanamori, 2004] A. Kanamori. Zermelo and Set Theory, Bulletin of Symbolic Logic, vol. 10, pp 487-553, 2004.
[Löwenheim, 1915] L. Löwenheim. Über Möglichkeiten im Relativkalkül, Math. Annalen, vol. 76, pp 447-470, 1915.
[Löwenheim, 1940] L. Löwenheim. Einkleidung der Mathematik in Schröderschen Relativkalkül, Journal of Symbolic Logic, vol. 5, pp 1-15, 1940.
[Mancosu *et al.*, 2005] P. Mancosu, R. Zach and C. Badesa. , The Development of Mathematical Logic from Russell to Tarski: 1900-1935, in L. Haaparanta, ed, The Development of Modern Logic, Oxford University Press, 2005.
[Mints, 1996] G. Mints. Thoralf Skolem and the Epsilon Substitution Method for Predicate Logic, Nordic Journal of Philosophical Logic, vol. 1, pp 133-146, 1996.
[Muller, 2004] F. Muller. Deflating Skolem, Synthese, vol 137, 2004.
[Post, 1947] E. Post. , Recursive unsolvability of a problem of Thue, Journal of Symbolic Logic, vol. 12, pp 1-11, 1947.
[Robinson, 1966] A. Robinson. Non-Standard Analysis, North Holland Publ. Comp., Amsterdam, 1966.
[Scott, 1960] D. Scott. On constructing models for arithmetic, in Infinistic Methods, Warsaw 1959, Pergamon Press, Oxford, pp 235-255, 1960.
[Thue, 1914] A. Thue. Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln, Vit.Ak. Skrifter 1914, Oslo, 1914.
[van den Dries and Levitz, 1984] L. van den Dries and H. Levitz. Trans. AMS, vol. 286, pp 339-349, 1984.
[van Heijenoort, 1967] J. van Heijenoort, ed. From Frege to Gödel. A Source Book in Mathematical Logic, 1879-1931, Harvard University Press, 1967.
[Vaught, 1974] R. Vaught. Model theory before 1945, in L. Henkin et al, eds, Proceedings of the Tarski Symposium, American Math. Society, Providence, pp 153-172, 1974.
[von Plato, 2003] J. von Plato. Skolem's Discovery of Gödel-Dummett Logic, Studia Logica, vol. 73, pp 153-157, 2003.
[von Plato, 2004] J. von Plato. Review of Kurt Gödel Collected Works, vol. IV-V, Bulletin of Symbolic Logic, pp 558-563, 2004.

[von Plato, 2007] J. von Plato. In the shadow of the Löwenheim–Skolem theorem: Early combinatorial analyses of the mathematical proofs, *Bulletin of Symbolic Logic*, 13, pp. 189-225, 2007.

[Wang, 1996] H. Wang. Skolem and Gödel, Nordic Journal of Philosophical Logic, vol. 1, pp 119-132, 1996.

[Weyl, 1910] H. Weyl. Über die Definitionen der mathematische Grundbegriffe, Mathematisch-naturwissenschaftliche Blätter, vol. 7, pp 93-95 and 109-113, 1910.

[Weyl, 1918] H. Weyl. Das Kontinuum, Veit, Leipzig, 1918.

[Whitehead and Russell, 1910–1913] A. N. Whitehead. and B. Russell. Principia Mathematica, Cambridge University Press, 1910-1913.

[Zermelo, 1929] E. Zermelo. Über den Begriff der Definitheit in der Axiomatik, Fund. Math., vol. 14, pp 339-344, 1929.

[Zermelo, 1930] E. Zermelo. Über Grenzzahlen und Mengenbereiche, Fund. Math., vol. 16, pp 29-47, 1930.

JACQUES HERBRAND: LIFE, LOGIC, AND AUTOMATED DEDUCTION*

Claus-Peter Wirth, Jörg Siekmann, Christoph Benzmüller, Serge Autexier

1 PREFACE

The lives of mathematical prodigies who passed away very early after groundbreaking work invoke a fascination for later generations: The early death of Niels Henrik Abel (1802–1829) from ill health after a sled trip to visit his fiancé for Christmas; the obscure circumstances of Evariste Galois' (1811–1832) duel; the deaths of consumption of Gotthold Eisenstein (1823–1852) (who sometimes lectured his few students from his bedside) and of Gustav Roch (1839–1866) in Venice; the drowning of the topologist Pavel Samuilovich Urysohn (1898–1924) on vacation; the burial of Raymond Paley (1907–1933) in an avalanche at Deception Pass in the Rocky Mountains; as well as the fatal imprisonment of Gerhard Gentzen (1909–1945) in Prague[1] — these are tales most scholars of logic and mathematics have heard in their student days.

Jacques Herbrand, a young prodigy admitted to the École Normale Supérieure as the best student of the year 1925, when he was 17, died only six years later in a mountaineering accident in La Bérarde (Isère) in France. He left a legacy in logic and mathematics that is outstanding.

Inevitably, when all introductory words are said in the annual postgraduate course on logic and automated deduction, the professor will feel the urge to point out to the young students that there are things beyond the latest developments of computer technology or the fabric of the Internet: eternal truths valid on planet Earth but in all those far away galaxies just as well.

And as there are still ten minutes to go till the end of the lecture, the students listen in surprise to the strange tale about the unknown flying objects from the far away, now visiting planet Earth and being welcomed by a party of human dignitaries from all strata of society. Not knowing what to make of all this, the little green visitors will ponder the state of evolution on this strange but beautiful planet: obviously life is there — but can it think?

*Second Readers: Irving H. Anellis, Paolo Bussotti, Lawrence C. Paulson.

[1]Cf. [Vihan, 1995].

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

The Earthlings seem to have flying machines, they are all connected planet-wide by communicators — but can they really think? Their gadgets and little pieces of machinery appear impressive — but is there a true civilization on planet Earth? How dangerous are they, these Earthlings made of a strange protoplasmic variant of matter?

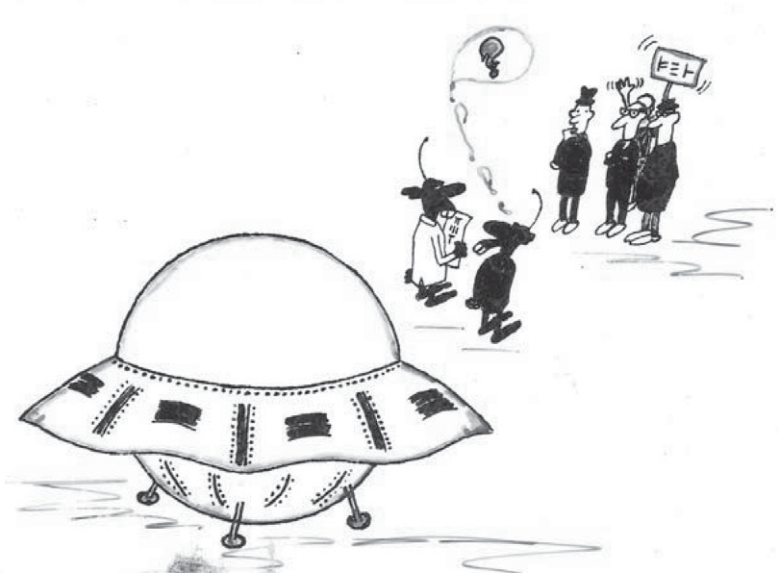
And then cautiously looking through the electronic windows of their flying unknown objects, they notice that strange little bearded Earthling, being pushed into the back by the more powerful dignitaries, who holds up a sign post with

$$\boxed{\models \equiv \vdash}$$

written on it. Blank faces, not knowing what to make of all this, the oldest and wisest scientist is slowly moved out through the e-door of the flying object, slowly being put down to the ground, and now the bearded Earthling is asked to come forward and the two begin that cosmic debate about syntax and semantics, proof theory and model theory, while the dignitaries stay stunned and silent. And soon there is a sudden flash of recognition and a warm smile on that green and wrinkled old face, who has seen it all and now waves back to his fellow travelers who remained safely within the flying object: "Yes, they have minds — yes oh yes!"

And this is why the name "Jacques Herbrand" is finally written among others with a piece of chalk onto the blackboard — and now that the introductory lecture is coming to a close, the professor promises to tell next week, what this name stands for and what that young scientist found out when he was only 21 years old.

And when the professor had grown old and frail, having told that tale more often than he cared to remember, the students presented him with a framed little birthday drawing that looked almost like this:



2 LIFE

This brief resumé of Jacques Herbrand's life focuses on his entourage and the people he met.[2] He was born on Feb. 12, 1908, in Paris, France, where his father, Jacques Herbrand Sr., worked as a trader in antique paintings.[3] He remained the only child of his parents, who were of Belgian origin. He died — 23 years old — in a mountaineering accident on July 27, 1931, in La Bérarde, Isère, France.

In 1925, only 17 years old, he was ranked first at the entrance examination to the prestigious *École Normale Supérieure* (*ENS*)[4] in Paris — but he showed little interest for the standard courses at the Sorbonne, which he considered a waste of time.[3] However, he closely followed the famous "Séminaire Hadamard" at the Collège de France, organized by Jacques Salomon Hadamard (1865–1963), from 1913 until 1933.[5] That seminar attracted many students. At Herbrand's time, among these students, prominent in their later lives, were:

name	lifetime	year of entering the ENS
André Weil	(1906–1998)	1922
Jean Dieudonné	(1906–1992)	1924
Jacques Herbrand	(1908–1931)	1925
Albert Lautman	(1908–1944)	1926
Claude Chevalley	(1909–1984)	1926

Weil, Dieudonné, and Chevalley would later be known among the founders of the renowned Bourbaki group: the French mathematicians who published the book series on the formalization of mathematics.

Weil, Lautman, and Chevalley were Herbrand's friends. Chevalley and Herbrand became particularly close friends[6] and they worked together on algebra.[7] Chevalley depicts Herbrand as an adventurous, passionate, and often perfectionistic personality who was not only interested in mathematics, but also in poetry and sports.[3] In particular, he seemed to have liked extreme sportive challenges: mountaineering, hiking, and long distance swimming. His interest in philosophical issues and foundational problems of science was developed well beyond his age.

At that time, the ENS did not award a diploma, but the students had to prepare the *agrégation*, an examination necessary to be promoted to *professeur agrégé*,[8] even though most students engaged into research. Herbrand passed the agrégation

[2] More complete accounts of Herbrand's life and personality can be found in [Chevalley, 1935; 1982], [Chevalley and Lautman, 1931], [Gödel, 1986ff., Vol. V, pp. 3–25]. All in all, very little is known about his personality and life.

[3] Cf. [Chevalley, 1982].

[4] Cf. `http://www.ens.fr` for the ENS in general and `http://www.archicubes.ens.fr` for the former students of the ENS.

[5] Cf. [Cartwright, 1965, p. 82], [Dieudonné, 1999, p. 107].

[6] Catherine Chevalley, the daughter of Claude Chevalley has written to Peter Roquette on Herbrand: "he was maybe my father's dearest friend" [Roquette, 2000, p. 36, Note 44].

[7] Cf. [Herbrand and Chevalley, 1931].

[8] This corresponds to a high-school teacher. The original rôle of the ENS was to educate students to become high-school teachers. Also Jean-Paul Sartre started his career like this.

in 1928, again ranked first, and he prepared his doctoral thesis under the direction of Ernest Vessiot (1865–1952), who was the director of the ENS since 1927.[9]

He submitted his thesis [Herbrand, 1930], entitled *Recherches sur la théorie de la démonstration*, on April 14, 1929. It was approved for publication on June 20, 1929. In October that year he had to join the army for his military service which lasted for one year in those days. He finally defended his thesis on June 11, 1930.[10]

After completing his military service in September 1930, awarded with a Rockefeller Scholarship, he spent the academic year 1930–1931 in Germany and intended to stay in Princeton[11] for the year after. He visited the following mathematicians:

hosting scientist	lifetime	place	time of Herbrand's stay
John von Neumann	(1903–1957)	Berlin	{ middle of Oct. 1930
Emil Artin	(1898–1962)	Hamburg	{ middle of May 1931
Emmy Noether[12]	(1882–1935)	Göttingen	{ middle of June 1931
			{ middle of July 1931[13]

Herbrand discussed his ideas with Paul Bernays (1888–1977) in Berlin, and he met Paul Bernays, David Hilbert (1862–1943), and Richard Courant (1988–1972) later in Göttingen.[14] On April 7, 1931, Herbrand wrote a letter to Kurt Gödel (1906–1978), who answered with some delay on July 25, most probably too late for the letter to reach Herbrand before his early death two days later.[15]

In other words, although Jacques Herbrand was still a relatively unknown young scientist, he was well connected to the best mathematicians and logicians of his time, particularly to those interested in the foundations of mathematics.

Herbrand met Helmut Hasse (1898–1979) at the Schiefkörper-Kongress (Feb. 26 – March 1, 1931) in Marburg and later wrote several letters including plenty of mathematical ideas to Hasse.[16] After exchanging several compassionate letters with Herbrand's father,[17] Hasse wrote Herbrand's obituary which is printed as the foreword to Herbrand's article on the consistency of arithmetic [Herbrand, 1932a].

[9] It is interesting to note that Ernest Vessiot and Jacques Salomon Hadamard where the two top-ranked students at the examination for the ENS in 1884.

[10] One reason for the late defense is that because of the minor rôle mathematical logic played at that time in France, Herbrand's supervisor Ernest Vessiot had problems finding examiners for the thesis. The final committee consisted of Ernest Vessiot, Arnaud Denjoy (1884–1974) and Maurice René Fréchet (1878–1973).

[11] Cf. [Gödel, 1986ff., Vol. V, p. 3f.]. Cf. [Roquette, 2008] for the letters between Hasse, Herbrand, and Joseph H. M. Wedderburn (1882–1948) (Princeton) on Herbrand's visit to Princeton.

[12] Herbrand had met Noether already very early in 1931 in Halle an der Saale. Cf. [Lemmermeyer and Roquette, 2006, p.106, Note 10].

[13] According to [Dubreil, 1983, p.73], Herbrand stayed in Göttingen until the *beginning* of July 1931. In his report to the Rockefeller Foundation Herbrand wrote that his stay in Germany lasted from Oct. 20, 1930, until the *end* of July 1931, which is unlikely because he died on July 27, 1931, in France.

[14] That Herbrand met Bernays, Hilbert, and Courant in Göttingen is most likely, but we cannot document it. Hilbert was still lecturing regularly in 1931, cf. [Reid, 1970, p.199]. Courant wrote a letter to Herbrand's father, cf. [Herbrand, 1971, p. 25, Note 1].

[15] Cf. [Gödel, 1986ff., Vol. V, pp. 3–25].

[16] Cf. [Roquette, 2007; 2008].

[17] Cf. [Hasse and Herbrand Sr., 1931].

3 LOGIC

3.1 Finitistic Classical First-Order Proof Theory

Herbrand's work on logic falls into the area of what is called *proof theory* today. More specifically, he is concerned with the *finitistic* analysis of *two-valued* (i.e. *classical*), *first-order* logic and its relationship to sentential, i.e. propositional logic.

Over the millenia, logic developed as *proof theory*. The key observation of the ancient Greek schools, first formulated by Aristotle of Stagira (384–322 B.C.), is that certain patterns of reasoning are valid irrespective of their actual denotation. From "all men are mortal" and "Socrates is a man" we can conclude that "Socrates is mortal", irrespectively of Socrates' most interesting personality and the contentious meaning of "being mortal" in this and other possible worlds. The discovery of those patterns of reasoning, called *syllogisms*, where meaningless symbols are used instead of everyday words, was the starting point of the known history of mathematical logic in the ancient world. For over two millenia, the development of these rules for drawing conclusions from given assumptions just on the basis of their *syntactical form* was the main subject of logic.

Model theory — on the other hand — the study of formal languages and their interpretation, became a respectable way of logical reasoning through the seminal works of Leopold Löwenheim (1878–1957) and Alfred Tarski (1901–1983). Accepting the actual infinite, model theory considers the set-theoretic semantical structures of a given language. With Tarski's work, the relationship between these two areas of logic assumed overwhelming importance — as captured in our little anecdote in the preface where '$\models$' signifies model-theoretic validity and '$\vdash$' denotes proof-theoretic derivability.[18]

Herbrand's scientific work coincided with the maturation of modern logic, as marked inter alia by Gödel's incompleteness theorems of 1930–1931.[19] It was strongly influenced by the foundational crisis in mathematics as well. Russell's Paradox was not only a personal calamity to Gottlob Frege (1848–1925),[20] but it jeopardized the whole enterprise of set theory and thus the foundation of modern mathematics. From an epistemological point of view, maybe there was less reason for getting scared as it appeared at the time: As Wittgenstein (1889–1951) reasoned later,[21] the detection of inconsistencies is an inevitable element of human learning, and many logicians today would be happy to live at such an interesting time of foundational crisis.

[18] As we will discuss in § 3.12, Jacques Herbrand still had problems in telling '$\vdash$' and '$\models$' apart: For instance, he blamed Löwenheim for not showing consistency of first-order logic, which is a property related to Herbrand's '$\vdash$', but not to Löwenheim's '$\models$'.

[19] Cf. [Gödel, 1931], [Rosser, 1936]. For an interesting discussion of the reception of the incompleteness theorems cf. [Dawson Jr., 1991].

[20] Cf. [Frege, 1893/1903, Vol. II].

[21] Cf. [Wittgenstein, 1939].

The development of mathematics, however, more often than not attracts the intelligent young men looking for clarity and reliability in a puzzling world, threatened by social complexity. As David Hilbert put it:

> "Und wo sonst soll Sicherheit und Wahrheit zu finden sein, wenn sogar das mathematische Denken versagt?"[22]

> "Es bildet ja gerade einen Hauptreiz bei der Beschäftigung mit einem mathematischen Problem, dass wir in uns den steten Zuruf hören: da ist ein Problem, suche die Lösung; du kannst sie durch reines Denken finden; denn in der Mathematik gibt es kein Ignorabimus."[23]

Furthermore, Hilbert did not want to surrender to the new "intuitionistic" movements of Luitzen Brouwer (1881–1966) and Hermann Weyl (1885–1955), who suggested a restructuring of mathematics with emphasis on the problems of existence and consistency rather than elegance, giving up many previous achievements, especially in analysis and in the set theory of Georg Cantor (1845–1918):[24]

> "Aus dem Paradies, das uns Cantor geschaffen, soll uns niemand vertreiben."[25]

Building on the works of Richard Dedekind (1831–1916), Charles S. Peirce (1839–1914), Ernst Schröder (1841–1902), Gottlob Frege (1848–1925), and Guiseppe Peano (1858–1932), the celebrated three volumes of *Principia Mathematica* [Whitehead and Russell, 1910–1913] of Alfred North Whitehead (1861–1947) and Bertrand Russell (1872–1970) had provided evidence that — in principle — mathematical proofs could be reduced to logic, using only a few rules of inference and appropriate axioms.

The goals of *Hilbert's programme* on the foundation of mathematics, however, extended well beyond this: His contention was that the reduction of mathematics to formal theories of logical calculi would be insufficient to resolve the foundational crisis of mathematics, neither would it protect against Russell's Paradox and other inconsistencies in the future — unless the consistency of these theories could be shown formally by "simple means".

---

[22] Cf. [Hilbert, 1926, p.170].
"And where else are security and truth to be found, if even mathematical thinking fails?"

[23] Cf. [Hilbert, 1926, p.180, modernized orthography].
"After all, one of the things that attract us most when we apply ourselves to a mathematical problem is precisely that within us we always hear the call: here is a problem, search for the solution; you can find it by pure thought, for in mathematics there is no *ignorabimus*."
[Heijenoort, 1971a, p. 384, translation by Stefan Bauer-Mengelberg]

[24] Cf. [Brouwer, 1925a; 1925b; 1926], [Weyl, 1921; 1928], [Cantor, 1932].

[25] Cf. [Hilbert, 1926, p.170].
"No one shall drive us from the paradise Cantor has created."

Hilbert's view on mathematics in the 1920s was influenced by Kant's *Critique of pure reason.*[26] Mathematics as directly and intuitionally perceived by a mathematician is called *contentual* (*inhaltlich*) by Hilbert. According to [Hilbert, 1926], the notions and methods of contentual mathematics are partly abstracted from finite symbolic structures (such as explicitly and concretely given natural numbers, proofs, and algorithms) where we can effectively *decide* (in finitely many effective steps) whether a given object has a certain property or not. Beyond these *aposterioristic* abstractions from phenomena, contentual mathematics also has a *synthetic*[27] *aprioristic*[28] aspect, which depends neither on experience nor on deduction, and which cannot be reduced to logic. Or, as Hilbert put it:

> "Schon Kant hat gelehrt – und zwar bildet dies einen integrierenden Bestandteil seiner Lehre –, dass die Mathematik über einen unabhängig von aller Logik gesicherten Inhalt verfügt und daher nie und nimmer allein durch Logik begründet werden kann, weshalb auch die Bestrebungen von Frege und Dedekind scheitern mussten. Vielmehr ist als Vorbedingung für die Anwendung logischer Schlüsse und für die Betätigung logischer Operationen schon etwas in der Vorstellung gegeben: gewisse, außer-logische konkrete Objekte, die anschaulich als unmittelbares Erlebnis vor allem Denken da sind."[29]

[26] The *Critique of pure reason* [Kant, 1787] elaborates how little we can know about things independent of an observer (*things in themselves*, *Dinge an sich selbst*) in comparison to our conceptions, i.e. the representations of the things within our thinking (Erscheinungen, Vorstellungen, Anschauungen). In what he compared to the Kopernikan revolution [Kant, 1787, p. XVI], Kant considered the conceptions gained in connection with sensual experience to be real and partly objectifiable, and accepted the things in themselves only as limits of our thinking, about which nothing can be known for certain.

[27] "synthetic" is the opposite of "analytic" and means that a statement provides new information that cannot be deduced from a given knowledge base. Contrary to Kant's opinion that all mathematical theorems are synthetic [Kant, 1787, p.14], (contentual) mathematics also has analytic sentences. In particular, Kant's example $7+5=12$ becomes analytic when we read it as $\mathsf{s}^7(\mathsf{0})+\mathsf{s}^5(\mathsf{0})=\mathsf{s}^{12}(\mathsf{0})$ and assume the non-necessary, synthetic, aprioristic axioms $x+\mathsf{0}=x$ and $x+\mathsf{s}(y)=\mathsf{s}(x+y)$. Cf. [Frege, 1884, § 89].

[28] "*a priori*" is the opposite of "*a posteriori*" and means that a statement does not depend on any form of experience. For instance, all necessary [Kant, 1787, p. 3] and all analytic [Kant, 1787, p.11] statements are *a priori*. Finally, Kant additionally assumes that all aprioristic statements are necessary [Kant, 1787, p. 219], which seems to be wrong, cf. Note 27.

[29] Cf. [Hilbert, 1926, p.170f., modernized orthography].
"Kant already taught—and indeed it is part and parcel of his doctrine—that mathematics has at its disposal a content secured independently of all logic and hence can never be provided with a foundation by means of logic alone; that is why the efforts of Frege and Dedekind were bound to fail. Rather, as a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation, certain extra-logical concrete objects that are intuitively present as immediate conceptions prior to all thought." [Heijenoort, 1971a, p. 376, translation by Stefan Bauer-Mengelberg, modified[30]]

[30] We have replaced "extralogical" with "extra-logical", and — more importantly — "experience" with "conception", for the following reason: Contrary to "Erfahrung" (experience), the German word "Erlebnis" does not suggest an *aposterioristic* intention, which would contradict the obviously *aprioristic* intention of Hilbert's sentence.

Kant and Hilbert follow Plato in using the word "ideal" to refer to those *things in themselves* (*Dinge an sich selbst*) which are not material, but intellectual concepts. *Ideal* objects and methods in mathematics (as opposed to contentual ones) may involve the actual infinite, such as quantification, ε-binding, set theory, and non-terminating computations. According to both [Kant, 1787] and [Hilbert, 1926], the only possible criteria for the acceptance of *ideal* notions are *consistency* and *usefulness*. Contrary to Kant,[31] however, Hilbert is willing to accept useful ideal theories, provided that their consistency can be shown with contentual and intuitively clear methods — i.e. "simple means".

These "simple means" that may be admitted here must be, on the one hand, sufficiently expressive and powerful to show the consistency of arithmetic, but, on the other hand, simple, i.e. intuitively clear and contentually reliable. The notion of *Hilbert's finitism* was born out of the conflict of these two goals.

Moreover, Hilbert expresses the hope that the new proof theory, primarily developed to show the consistency of ideal mathematics with contentual means, would also admit (possibly ideal, i.e. non-finitistic) proofs of *completeness*[32] for certain mathematical theories. If this goal of Hilbert's programme had been achieved, then ideal proofs would have been justified as convenient short-cuts for constructive, contentual, and intuitively clear proofs, so that — even under the threat of Russell's Paradox and others — there would be no reason to give up the paradise of axiomatic mathematics and abstract set theory.

And these basic convictions of the Hilbert school constituted the most important influence on the young student of mathematics Jacques Herbrand.

As Gödel showed with his famous incompleteness theorems in 1930–1931,[19] however, the consistency of any (reasonably conceivable) recursively enumerable mathematical theory that includes arithmetic excludes both its completeness and the existence of a finitistic consistency proof.

Nevertheless, the contributions of Wilhelm Ackermann (1896–1962), Bernays, Herbrand, and Gentzen within Hilbert's programme gave *proof theory* a new meaning as a field in which proofs are the objects and their properties and constructive transformations are the field of mathematical study, just as in arithmetic the numbers are the objects and their properties and algorithms are the field of study.

[31] Kant considers ideal notions to be problematic, because they transcend what he considers to be the area of objective reality; cf. Note 26. For notions that are consistent, useful, and ideal, Kant actually introduces the technical term *problematic* (problematisch):

> "Ich nenne einen Begriff problematisch, der keinen Widerspruch enthält, der auch als eine Begrenzung gegebener Begriffe mit anderen Erkenntnissen zusammenhängt, dessen objektive Realität aber auf keine Weise erkannt werden kann."
>
> [Kant, 1787, p. 310, modernized orthography]

[32] A theory is *complete* if for any formula A without free variables (i.e. any closed formula in the given language) which is not part of this theory, its negation $\neg A$ is part of this theory.

3.2 Herbrand's Contributions and his Notion of Intuitionism

The essential works of Herbrand on logic are his Ph.D. thesis [Herbrand, 1930] and the subsequent journal article [Herbrand, 1932a], both to be found in [Herbrand, 1971].[33]

The main contribution is captured in what is called today *Herbrand's Fundamental Theorem*. Sometimes it is simply called "Herbrand's Theorem", but the longer name is preferable as there are other important "Herbrand theorems", such as the Herbrand–Ribet Theorem. Moreover, Herbrand himself calls it "Théorème fondamental".

The subject of Herbrand's Fundamental Theorem is the effective reduction of (the semi-decision problem of) provability in first-order logic to provability in sentential logic.

Here we use the distinction well-known to Herbrand and his contemporaries between *first-order logic* (where quantifiers bind variables ranging over objects of the universe, i.e. of the domain of reasoning or discourse) and *sentential logic* without any quantifiers. Validity of a formula in sentential logic is effectively decidable, for instance with the truth-table method.

Although Herbrand spends Chapter 1 of his thesis on the subject, he actually "shows no interest for the sentential work" [Heijenoort, 1986c, p.120], and takes it for granted.

A. Contrary to Gentzen's Hauptsatz [Gentzen, 1935], Herbrand's Fundamental Theorem starts right with a *single sentential tautology* (cf. § 3.7). He treats this property as given and does not fix a concrete method for establishing it.

B. The way Herbrand presents his sentential logic in terms of '$\neg$' and '$\vee$' indicates that he is not concerned with *intuitionistic logic* as we understand the term today.[34]

Contrary to Gentzen's sequent calculus LK [Gentzen, 1935], in Herbrand's calculi we do not find something like a sub-calculus LJ for intuitionistic logic. Moreover, there is no way to generalize Herbrand's Fundamental Theorem to include intuitionistic logic: Contrary to the Cut elimination in Gentzen's Hauptsatz, the elimination of *modus ponens* according to Herbrand's Fundamental Theorem does not hold for intuitionistic logic.

[33] This book is still the best source on Herbrand's writings today. It is not just an English translation of Herbrand's complete works on logic (all of Herbrand's work in logic was written in French, cf. [Herbrand, 1968]), but contains additional annotation, brief introductions, and extended notes by Jean van Heijenoort, Burton Dreben, and Warren Goldfarb. Besides minor translational corrections, possible addenda for future editions would be the original texts in French; Herbrand's mathematical writings outside of logic; some remarks on the two corrections of Herbrand's False Lemma by Gödel and Heijenoort, respectively, cf. §§ 3.9 and 3.10 below; and Herbrand's correspondence. The correspondence with Gödel is published in [Gödel, 1986ff., Vol. V, pp. 14–25]. Herbrand's letters to Hasse are still in private possession according to [Roquette, 2007]. The whereabouts of the rest of his correspondence is unknown.

[34] Cf. e.g. [Heyting, 1930a; 1971], [Gentzen, 1935].

> "All the attempts to generalize Herbrand's theorem in that direction have only led to partial and unhandy results (see [Heijenoort, 1971b])."
> [Heijenoort, 1986c, p.120f.]

When Herbrand uses the term "intuitionism", this typically should be understood as referring to something closer to the finitism of Hilbert than to the intuitionism of Brouwer.[35] This ambiguous usage of the term "intuitionism" — centered around the partial rejection of the Law of the Excluded Middle, the actual infinite, as well as quantifiers and other binders — was common in the Hilbert school at Herbrand's time.[36] Herbrand's view on what he calls "intuitionism" is best captured in the following quote:

> "Nous entendons par raisonnement intuitionniste, un raisonnement qui satisfait aux conditions suivantes: on n'y considère jamais qu'un nombre fini déterminé d'objets et de fonctions; celles-ci sont bien définies, leur définition permettant de calculer leur valeur d'une manière univoque; on n'affirme jamais l'existence d'un objet sans donner le moyen de le construire; on n'y considère jamais l'ensemble de tous les objets x d'une collection infinie; et quand on dit qu'un raisonnement (ou un théorème) est vrai pour tous ces x, cela signifie que pour chaque x pris en particulier, il est possible de répéter le raisonnement général en question qui ne doit être considéré que comme le prototype de ces raisonnements particuliers."[37]

Contrary to today's precise meaning of the term "intuitionistic logic", the terms "intuitionism" and "finitism" denote slightly different concepts, which are related to the philosophical background, differ from person to person, and vary over times.[38]

[35] Gödel, however, expressed a different opinion in a letter to Heijenoort of Sept. 18, 1964:

> "In Note 3 of [Herbrand, 1932a] he does *not* require the enumerability of mathematical objects, and gives a definition which fits Brouwer's intuitionism very well"
> [Gödel, 1986ff., Vol. V, p. 319f.]

[36] Cf. e.g. [Herbrand, 1971, p. 283f.], [Tait, 2006, p. 82ff.].

[37] Cf. [Herbrand, 1932a, p. 3, Note 3]. Without the comma after "intuitionniste", but with an additional comma after "question" also in: [Herbrand, 1968, p. 225, Note 3].
"By an intuitionistic argument we understand an argument satisfying the following conditions: in it we never consider anything but a given finite number of objects and of functions; these functions are well-defined, their definition allowing the computation of their value in a univocal way; we never state that an object exists without giving the means of constructing it; we never consider the totality of all the objects x of an infinite collection; and when we say that an argument (or a theorem) is true for all these x, we mean that, for each x taken by itself, it is possible to repeat the general argument in question, which should be considered to be merely the prototype of the particular arguments." [Herbrand, 1971, Note 5, p. 288f., translation by Heijenoort]

[38] Regarding intuitionism, besides Brouwer, Weyl, and Hilbert, we may count Leopold Kronecker (1823–1891) and Henri Poincaré (1854–1912) among the ancestors, and have to mention Arend Heyting (1898–1980) for his major differing view, cf. e.g. [Heyting, 1930a; 1930b; 1971]. Deeper discussions of Herbrand's notion of "intuitionism" can be found in [Heijenoort, 1986c, pp. 113–118] and in [Tait, 2006, p. 82ff.]. Moreover, we briefly discuss it in Note 123. For more on finitism cf. e.g. [Parsons, 1998], [Tait, 1981], [Zach, 2001]. For more on Herbrand's background in philosophy of mathematics, cf. [Chevalley, 1935], [Dubucs and Égré, 2006].

While Herbrand is not concerned with intuitionistic logic, he is a finitist with respect to the following two aspects:

1. Herbrand's work is strictly contained within Hilbert's finitistic programme and he puts ample emphasis on his finitistic standpoint and the finitistic character of his theorems.
2. Herbrand does not accept any model-theoretic semantics unless the models are finite. In this respect, Herbrand is more finitistic than Hilbert, who demanded finitism only for consistency proofs.

> "Herbrand's negative view of set theory leads him to take, on certain questions, a stricter attitude than Hilbert and his collaborators. He is more royalist than the king. Hilbert's metamathematics has as its main goal to establish the consistency of certain branches of mathematics and thus to justify them; there, one had to restrict himself to finitistic methods. But in logical investigations other than the consistency problem of mathematical theories the Hilbert school was ready to work with set-theoretic notions."
>
> [Heijenoort, 1986c, p.118]

3.3 The Context of Herbrand's Work

Let us now have a look at what was known in Herbrand's time and at the papers that influenced his work on logic.

Stanisław Zaremba (1863–1942) is mentioned in [Herbrand, 1928], where he cites Zaremba's textbook on mathematical logic [Zaremba, 1926], which clearly influenced Herbrand's notation.[39]

Herbrand's subject, first-order logic, became a field of special interest not least because of the seminal paper [Löwenheim, 1915], which singled out first-order logic in the *Theory of Relatives* developed by Peirce and Schröder.[40] With this paper, first-order logic became an area of special interest, due to the surprising meta-mathematical properties of this logic, which was intended to be an especially useful tool with a restricted area of application.[41]

As the presentation in [Löwenheim, 1915] is opaque, Thoralf Skolem (1887–1963) wrote five clarifying papers contributing to the substance of Löwenheim's Satz 2, the now famous Löwenheim–Skolem Theorem [Skolem, 1920; 1923b; 1928;

[39] Cf. Goldfarb's Note to [Herbrand, 1928] on p. 32ff. in [Herbrand, 1971]. Zaremba was one of the leading Polish mathematicians in the 1920s. He had close connections to Paris, but we do not know whether Herbrand ever met him.

[40] For the heritage of Peirce cf. [Peirce, 1885], [Brady, 2000]; for that of Schröder cf. [Schröder, 1895], [Brady, 2000].

[41] Without set theory, first-order logic was too poor to serve as such a single universal logic as the ones for which Frege, Peano, and Russell had been searching; cf. [Heijenoort, 1986a]. For the suggestion of first-order logic as the basis for set theory, we should mention [Skolem, 1923b], which is sometimes cited as of the year 1922, and therefore easily confused with [Skolem, 1923a]. For the emergence of first-order logic as the basis for mathematics see [Moore, 1987].

1929; 1941]. From these papers, Herbrand cites [Löwenheim, 1915] and [Skolem, 1920], and the controversy pro and contra Herbrand's reading of [Skolem, 1923b] and [Skolem, 1928] will be presented in § 3.12 below.

While Herbrand neither cites Peano nor mentions Frege, the *Principia Mathematica* [Whitehead and Russell, 1910–1913] were influential at his time, and he was well aware of it. Herbrand cites all editions of the *Principia* and there are indications that he studied parts of it carefully.[42] But Russell's influence is comparatively minor compared to Hilbert's, as, indeed, Herbrand was most interested in proving consistency, decidability, and completeness. Jean van Heijenoort (1912–1986) notes on Herbrand:

> "The difficulties provoked by the Russell Paradox, stratification, ramification, the problems connected with the axiom of infinity or the axiom of reducibility, nothing of that seems to retain his attention.
> The reason for this attitude is that Herbrand does not share Russell's conception concerning the relation between logic and mathematics, but had adopted Hilbert's. In 1930 Herbrand indicates quite well where he sees the limits of Russell's accomplishment: 'So far we have only replaced ordinary language with another more convenient one, but this does not help us at all with respect to the problems regarding the principles of mathematics.' [Herbrand, 1930a, p. 248; 1971, p. 208]. And the sentence that follows indicates the way to be followed: 'Hilbert sought to resolve the questions which can be raised by applying himself to the study of collections of signs which are translations of propositions true in a determinate theory.' "[43]

As Herbrand's major orientation was toward the Hilbert school, it is not surprising that the majority of his citations[44] refer to mathematicians related either to the Hilbert school or to Göttingen, which was the Mecca of mathematicians until its intellectual and organizational destruction by the Nazis in 1933.

By the title of his thesis *Recherches sur la théorie de la démonstration*, Herbrand clearly commits himself to King Hilbert's following, and being the first contributor to Hilbert's finitistic programme in France, he was given the opportunity to write on Hilbert's logic in a French review journal, and this paper [Herbrand, 1930a] is historically interesting because it captures Herbrand's personal view of Hilbert's finitistic programme.

[42] Herbrand seems to have studied ∗9 and ∗10 of [Whitehead and Russell, 1910–1913, Vol. I] carefully, leaving traces in Herbrand's Property A and in Chapter 2 of Herbrand's Ph.D. thesis, cf. [Heijenoort, 1986c, pp. 102–106].

[43] Cf. [Heijenoort, 1986c, p.105].

[44] In his thesis [Herbrand, 1930], Herbrand cites [Ackermann, 1925], [Artin and Schreier, 1927], [Behmann, 1922], [Bernays, 1928], [Bernays and Schönfinkel, 1928], [Hilbert, 1922; 1926; 1928], [Hilbert and Ackermann, 1928], and [Neumann, 1925; 1927; 1928]. Furthermore, Herbrand cites [Nicod, 1917] in [Herbrand, 1930], [Ackermann, 1928] in [Herbrand, 1931], and [Hilbert, 1931] and [Gödel, 1931] in [Herbrand, 1932a].

3.4 A Genius with some Flaws

On the one hand, Herbrand was a creative mathematician whose ideas were truly outstanding, not only for his time. Besides logic, he also contributed to class-field theory and to the theory of algebraic number fields. Although this is not our subject here, we should keep in mind that Herbrand's contributions to algebra are as important from a mathematical point of view and as numerous as his contributions to logic.[45] Among the many statements about Herbrand's abilities as a mathematician is Weil's letter to Hasse in August 1931 where he writes that he would not need to tell him what a loss Herbrand's death means especially for number theory.[46] As Peter B. Andrews put it: "Herbrand was by all accounts a brilliant mathematician."[47]

On the other hand, Herbrand neither had the education nor the supervision to present his results in proof theory with the technical rigor and standard, say, of the Hilbert school in the 1920s, let alone today's emphasis on formal precision.[48] Finitistic proof theory demands the disambiguation of form and content and a higher degree of precision than most other mathematical fields. Moreover, the field was novel at Herbrand's time and probably hardly anybody in France was able to advise Herbrand competently. Therefore, Herbrand, a *génie créateur*, as Heijenoort called him,[49] was apt to make undetected errors. Well known today is a serious flaw in his thesis which stayed unnoticed by its reviewers at the time. Moreover, several theorems are in fact conceptually correct, but incorrectly formulated.

Let us have a look at three flaws in § 3.3 of the Chapters 2, 3 and 5, respectively:

[45] Cf. [Herbrand, 1930b; 1931e; 1931f; 1932b; 1932c; 1932d; 1932e; 1932f; 1933; 1936] and [Herbrand and Chevalley, 1931], as well as [Dieudonné, 1982].

[46] Cf. [Lemmermeyer and Roquette, 2006, p.119, Note 6].

[47] Cf. [Andrews, 2003, p.171].

[48] Heijenoort writes in his well-known "source book":

> "Herbrand's thesis bears the marks of hasty writing; this is especially true of Chapter 5. Some sentences are poorly constructed, and the punctuation is haphazard. Herbrand's thoughts are not nebulous, but they are so hurriedly expressed that many a passage is ambiguous or obscure. To bring out the proper meaning of the text the translators had to depart from a literal rendering, and more rewriting has been allowed in this translation than in any other translation included in the present volume." [Heijenoort, 1971a, p. 525]

Similarly, Goldfarb, the translator and editor of Herbrand's logical writings, writes:

> "Herbrand also tended to express himself rather hastily, resulting in many obscurities; in these translations an attempt has been made to balance the demands of literalness and clarity." [Herbrand, 1971, p.V]

[49] Cf. [Herbrand, 1968, p.1].

Chapter 2, § 3.3: A typical instance for an incorrectly formulated theorem which is conceptually correct can be found in Chapter 2, § 3.3, on deep inference: *From* $\vdash B \Rightarrow C$ *we can conclude* $\vdash A[B] \Rightarrow A[C]$, *provided that* $[\cdots]$ *denotes only positive positions*[50] *in* A.
Herbrand, however, states

$$\vdash (B \Rightarrow C) \;\Rightarrow\; (A[B] \Rightarrow A[C])$$

which is not valid; to wit[51] apply the substitution

$$\{\ A[\ldots] \mapsto \forall x.\,[\ldots],\ \ B \mapsto \mathsf{true},\ \ C \mapsto \mathsf{P}(x)\ \}.$$

Chapter 3, § 3.3: Incorrectly formulated is also Herbrand's theorem on the relativization of quantifiers.[52] This error was recognized later by Herbrand himself.[53]

Moreover, notice that in this context, Herbrand discusses the *many-sorted* first-order logic related to the restriction to the language where all quantifiers are relativized, extending a similar discussion found already in [Peirce, 1885].

All of this is not terribly interesting, except that it gives us some clues on how Herbrand developed his theorems: It seems that he started, like any mathematician, with a strong intuition of the semantics and used it to formulate the theorem. Then he had a careful look at those parts of the proof that might violate the finitistic standpoint. The final critical check of minor details of the formalism in the actual proof,[54] however, hardly played a rôle in this work.

Chapter 5, § 3.3: The drawback of his intuitive style of work manifests itself in a serious mistake, which concerns a lemma that has crucial applications in the proof of the Fundamental Theorem, namely the "lemma" of Chapter 5, § 3.3, which we will call *Herbrand's False Lemma*. Before we discuss this in § 3.9, however, we have to define some notions.

3.5 Champs Finis, Herbrand Universe, and Herbrand Expansion

Most students of logic or computer science know Herbrand's name in the form of *Herbrand universe* or *Herbrand expansion*. Today, the *Herbrand universe* is usually defined as the set of all terms over a given signature, and the *Herbrand expansion* of a set of formulas results from a systematic replacement of all variables in that set of formulas with terms from the Herbrand universe.

[50] Note that a position in a formula (seen as a tree built-up from logical operators '$\wedge$', '$\vee$', '$\neg$', '$\forall$', and '$\exists$') is *positive* if the number of $\neg$-operators preceeding it on the path from the root position is even, and *negative* if it is uneven.

[51] Cf. also [Herbrand, 1971, Goldfarb's Notes 6 (p.78) and A (p. 98)].

[52] Relativization of quantifiers was first discussed in the elaboration of first-order logic in [Peirce, 1885]. Roughly speaking, it means to restrict quantifiers to a predicate P by replacing any $\forall x.\ A$ with $\forall x.\ (\mathsf{P}(x) \Rightarrow A)$, and, dually, $\exists x.\ A$ with $\exists x.\ (\mathsf{P}(x) \wedge A)$.

[53] In Note 1 of [Herbrand, 1932a].

[54] Cf. [Hadamard, 1949, Chapter V] for a nice account of this mode of mathematical creativity.

Historically, however, this is not quite correct. First of all, Herbrand does not use *term structures* for two reasons:

1. Herbrand typically equates terms with objects of the universe, and thereby avoids working explicitly with term structures.[55]
2. As a finitist more royal than King Hilbert, Herbrand does not accept structures with infinite universes.

As a finite substitute for a typically infinite full term universe, Herbrand uses what he calls a *champ fini* of order n, which we will denote with $\mathcal{T}_n$. Such a champ fini differs from a full term universe in containing only the terms t with $|t| < n$. We use $|t|$ to denote the *height* of the term t, which is given by

$$|f(t_1, \ldots, t_m)| \quad = \quad 1 + \max\{0, |t_1|, \ldots, |t_m|\}.$$

The terms of $\mathcal{T}_n$ are constructed from the function symbols and constant symbols (which we will tacitly subsume under the function symbols in the following) of a finite signature and from a finite set of variables. We will assume that an additional variable l, the *lexicon*, is included in this construction, if necessary to have $\mathcal{T}_n \neq \emptyset$.

Herbrand prefers to treat all logical symbols besides '$\neg$', '$\vee$', and '$\exists$' as defined. In prenex forms, the universal quantifier '$\forall$' is also treated as a primitive symbol.

Herbrand expansion was implicitly given already in the first elaborate description of first-order logic in [Peirce, 1885], under the name "first-intentional logic of relatives", following the invention of quantifiers in [Frege, 1879]. Herbrand speaks of "reduction" (*réduite*) instead of "Herbrand expansion".

DEFINITION 1 Herbrand Expansion, $A^{\mathcal{T}}$. For a finite set of terms $\mathcal{T}$, the expansion $A^{\mathcal{T}}$ of a formula A is defined as follows: $A^{\mathcal{T}} = A$ if A does not have a quantifier, $(\neg A)^{\mathcal{T}} = \neg A^{\mathcal{T}}$, $(A \vee B)^{\mathcal{T}} = A^{\mathcal{T}} \vee B^{\mathcal{T}}$, $(\exists x.\, A)^{\mathcal{T}} = \bigvee_{t \in \mathcal{T}} A^{\mathcal{T}}\{x \mapsto t\}$, and $(\forall x.\, A)^{\mathcal{T}} = \bigwedge_{t \in \mathcal{T}} A^{\mathcal{T}}\{x \mapsto t\}$, where $A^{\mathcal{T}}\{x \mapsto t\}$ denotes the result of applying the substitution $\{x \mapsto t\}$ to $A^{\mathcal{T}}$. □

EXAMPLE 2 Herbrand Expansion, $A^{\mathcal{T}}$.
For example, for $\mathcal{T} := \{\, 3,\ z+2 \,\}$, and for A being the arithmetic formula

$$\forall x.\, (x = 0 \quad \vee \quad \exists y.\, x = y + 1),$$

the expansion $A^{\mathcal{T}}$ is

$$\begin{pmatrix} & 3 = 0 \\ \vee & 3 = 3 + 1 \\ \vee & 3 = (z+2) + 1 \end{pmatrix} \quad \wedge \quad \begin{pmatrix} & z + 2 = 0 \\ \vee & z + 2 = 3 + 1 \\ \vee & z + 2 = (z+2) + 1 \end{pmatrix}.$$

□

The Herbrand expansion reduces a first-order formula to sentential logic in such a way that a sentential formula is reduced to itself, and that the semantics is invariant if the terms in $\mathcal{T}$ range over the whole universe. If, however, this is not the case — such as in our above example and for all infinite universes —, then Herbrand expansion changes the semantics by relativization of the quantifiers[52] to range only over those elements of the universe to which the elements of $\mathcal{T}$ evaluate.

[55] As this equating of terms has no essential function in Herbrand's works, but only adds extra complication to Herbrand's subjects, we will completely ignore it here and exclusively use free term structures in what follows.

3.6 Skolemization, Uniform Notation, γ- and δ-quantification

A first-order formula may contain existential as well as universal quantifiers. Can we make it more uniform by replacing either of them?

Consider the formula $\forall x.\, \exists y.\, \mathsf{Q}(x,y)$. These two quantifiers express a functional dependence between the values for x and y, which could also be expressed by a (new) function, say g, such that $\forall x.\, \mathsf{Q}(x,g(x))$, i.e. this function g chooses for each x the correct value for y, provided that it exists. In other words, we can replace any existentially quantified variable x which occurs in the scope of universal quantifiers for $y_1, \ldots, y_n$ with the new *Skolem term* $g(y_1,\ldots,y_n)$.

This replacement, carried out for all existential quantifiers, results in a formula having only universal quantifiers. Using the convention that all free variables are universally quantified, we may then just drop these quantifiers as well. Roughly speaking, this transformation, *Skolemization* as we call it today, leaves satisfiability invariant. It occurs for the first time explicitly in [Skolem, 1928], but was already used in an awkward formulation in [Schröder, 1895] and [Löwenheim, 1915].

For reasons that will become apparent later, Herbrand employs a form of Skolemization that is dual to the one above. Now the *universal* variables are removed first, so that all remaining variables are existentially quantified. How can this be done? Well, if the universally quantified variable x occurs in the scope of the existentially quantified variables $y_1,\ldots,y_m$, we can replace x with the *Skolem term* $x^\delta(y_1,\ldots,y_m)$. The second-order variable or first-order function symbol x^δ in this Skolem term stands for any function with arguments $y_1,\ldots,y_m$. Roughly speaking, this dual form of *Skolemization* leaves validity invariant.

For example, let us consider the formula $\exists y.\, \forall x.\, \mathsf{Q}(x,y)$. Assuming the axiom of choice and the standard interpretation of (higher-order) quantification, all of the following statements are logically equivalent:

- $\exists y.\, \forall x.\, \mathsf{Q}(x,y)$ holds.
- There is an object y such that $\mathsf{Q}(x,y)$ holds for every object x.
- $\exists y.\, \mathsf{Q}(x^\delta(y),y)$ holds for every function x^δ.
- $\forall f.\, \exists y.\, \mathsf{Q}(f(y),y)$ holds.

Now $\exists y.\, \mathsf{Q}(x^\delta(y),y)$ is called the *(validity) Skolemized form of* $\exists y.\, \forall x.\, \mathsf{Q}(x,y)$. The variable or function symbols x^δ of increased logical order are called *Skolem functions.*[56] The Skolemized form is also called *functional form* (with several addenda specifying the dualities), because Skolemization turns the object variable x into a function variable or function symbol $x^\delta(\cdots)$.

[56] Herbrand calls Skolem functions *index functions*, translated according to [Herbrand, 1971]. Moreover, in [Dreben and Denton, 1970], [Scanlon Jr., 1973], and [Goldfarb and Scanlon Jr., 1974], we find the term *indicial functions* instead of "index functions". The name "Skolem function" was used in [Gödel, 1939], probably for the first time, cf. [Anellis, 2006].

Note that $A \Rightarrow B$ and $\neg A \vee B$ and $\neg(A \wedge \neg B)$ are equivalent in two-valued logic. So are $\neg\forall x.\, A$ and $\exists x.\, \neg A$, as well as $\neg\exists x.\, A$ and $\forall x.\, \neg A$.

Accordingly, the *uniform notation* (as introduced in [Smullyan, 1968]) is a modern classification of formulas into only four categories: α, β, γ, and δ.

More important than the classification of formulas is the associated classification of the reductive inference rules applicable to them as *principal formulas*.

According to [Gentzen, 1935], but viewed under the aspect of reduction (i.e. the converse of deduction), the *principal formula* of an inference rule is the one which is (partly) replaced by its immediate "sub"-formulas, depending on its topmost operator.

- An α-formula is one whose validity reduces to the validity of a single operand of its topmost operator. For example, $A \vee B$ may be reduced either to A or to B, and $A \Rightarrow B$ may be reduced either to $\neg A$ or to B.
- A β-formula is one whose validity reduces to the validity of both operands of its topmost binary operator, introducing two cases of proof ($\beta = \underline{b}$ranching). For example, $A \wedge B$ reduces to both A and B, and $\neg(A \Rightarrow B)$ reduces to both A and $\neg B$.
- A γ-formula is one whose validity reduces to the validity of alternative instances of its topmost quantifier. For example, $\exists y.\, A$ reduces to $A\{y \mapsto y^\gamma\}$ in addition to $\exists y.\, A$, for a fresh *free γ-variable* y^γ. Similarly, $\neg\forall y.\, A$ reduces to $\neg A\{y \mapsto y^\gamma\}$ in addition to $\neg\forall y.\, A$. Free γ-variables may be globally instantiated at any time in a reduction proof.
- A δ-formula is one whose validity reduces to the validity of the instance of its topmost quantifier with its Skolem term. For example, $\forall x.\, A$ reduces to $A\{x \mapsto x^\delta(y_1^\gamma, \ldots, y_m^\gamma)\}$, where $y_1^\gamma, \ldots, y_m^\gamma$ are the free γ-variables already in use.[57]

For a more elaborate introduction into free γ- and δ-variables see [Wirth, 2008a].

Herbrand considers *validity* and *Skolemized form* as above in his thesis. In a similar context, which we will have to discuss below, Skolem considers *unsatisfiability*, a dual of validity, and *Skolem normal form* in addition to Skolemized form. As it was standard at his time, Herbrand called the two kinds of quantifiers — i.e. for γ- and δ-formulas[58] — *restricted* and *general* quantifiers, respectively. To avoid the problem of getting lost in several dualities in what follows, we prefer to speak of *γ-quantifiers* and *δ-quantifiers* instead. The variables bound by δ-quantifiers will be called *bound δ-variables*. The variables bound by γ-quantifiers will be called *bound γ-variables*.

[57] In the game-theoretic semantics of first-order logic, the δ-variables (such as x^δ in the above example) stand for the unknown choices by our opponent in the game, whereas, for showing validity, we have to specify a winning strategy by describing a finite number of first-order terms as alternative solutions for the γ-variables (such as y_i^γ above), cf. e.g. [Hintikka, 1996].

[58] Notice that it is obvious how to generalize the definition of α-, β-, γ- and δ-formulas from top positions to inner occurrences according to the category into which they would fall in a stepwise reduction. Therefore, we can speak of α-, β-, γ- and δ-formulas also for the case of subformulas and classify their quantifiers accordingly.

For a first-order formula A in which any bound variable is bound exactly once and does not occur again free, we define:

The *outer*[59] *Skolemized form of* A results from A by removing any δ-quantifier and replacing its bound variable x with $x^\delta(y_1, \ldots, y_m)$, where x^δ is a new symbol and $y_1, \ldots, y_m$, in this order,[60] are the variables of the γ-quantifiers in whose scope the δ-quantifier occurs.

3.7 Axioms and Rules of Inference

In the following we will present the calculi of Herbrand's thesis (i.e. the axioms and rules of inference) as required for our presentation of the Fundamental Theorem.

When we speak of a term, a formula, or a structure, we refer to first-order terminology without mentioning this explicitly. When we explicitly speak of "first order", however, it is to emphasize the contrast to sentential logic.

Sentential Tautology: Let B be a first-order formula. B is a *sentential tautology* if it is quantifier-free and truth-functionally valid, provided its atomic subformulas are read as atomic sentential variables.[61]

Modus Ponens: $\dfrac{A \qquad A \Rightarrow B}{B}$.

Generalized Rule of γ-Quantification: $\dfrac{A[B\{x \mapsto t\}]}{A[\gamma x.\, B]}$, where the free variables of the term t must not be bound by quantifiers in B, and γ stands for $\exists$ if $[\ldots]$ denotes a positive position[50] in $A[\ldots]$, and γ stands for $\forall$ if this position is negative. Moreover, we require that $[\ldots]$ does not occur in the scope of any quantifier in $A[\ldots]$. This requirement is not necessary for

[59] Herbrand has no name for the outer Skolemized form and he does not use the inner Skolemized form, which is the current standard in two-valued first-order logic and which is required for our discussion in Note 75. The *inner Skolemized form of* A results from A by repeating the following until all δ-quantifiers have been removed: Remove an outermost δ-quantifier and replace its bound variable x with $x^\delta(y_1, \ldots, y_m)$, where x^δ is a new symbol and $y_1, \ldots, y_m$, in this order, are the variables of the γ-quantifiers in whose scope the δ-quantifier occurs and which actually occur in the scope of the δ-quantifier. The inner Skolemized form is closely related to the liberalized δ-rule (also called δ^+-rule) in reductive calculi, such as sequent, tableau, or matrix calculi; cf. e.g. [Baaz and Fermüller, 1995], [Nonnengart, 1996], [Wirth, 2004, §§ 1.2.3 and 2.1.5], [Wirth, 2006], [Wirth, 2008a, § 4].

[60] Contrary to our fixation of the order of the variables as arguments to the Skolem functions (to achieve uniqueness of the notion), Herbrand does not care for the order in his definition of the outer Skolemized form. Whenever he takes the order into account, however, he orders by occurrence from left to right or else by the height of the terms w.r.t. a substitution, but not by the names of the variables.

[61] Notice that this notion is more restrictive than the following, which is only initially used by Herbrand in his thesis, but which is standard for the predicate calculi of the Hilbert school and the Principia Mathematica; cf. [Hilbert and Bernays, 1968/70, Vol. II, Supplement I D], [Whitehead and Russell, 1910–1913, *10]. B is a *substitutional sentential tautology* if there is a truth-functionally valid sentential formula A and a substitution σ mapping any sentential variable in A to a first-order formula such that B is $A\sigma$. For example, both $\mathsf{P}(x) \vee \neg\mathsf{P}(x)$ and $\exists x.\, \mathsf{P}(x) \vee \neg\exists x.\, \mathsf{P}(x)$ are substitutional sentential tautologies, related to the truth-functionally valid sentential formula $p \vee \neg p$, but only the first one is a sentential tautology.

soundness, but for the constructions in the proof of Herbrand's Fundamental Theorem.

E.g., with the meta-level substitution $\{\ A[...] \mapsto (t \prec t) \vee [...],\quad B \mapsto \neg(x \prec t)\ \}$ we get $\frac{(t \prec t) \vee \neg(t \prec t)}{(t \prec t) \vee \exists x.\ \neg(x \prec t)}$, and with $\{\ A[...] \mapsto (t \prec t) \vee \neg[...],\quad B \mapsto (x \prec t)\ \}$ we get $\frac{(t \prec t) \vee \neg(t \prec t)}{(t \prec t) \vee \neg\forall x.\ (x \prec t)}$.

Note that Herbrand considers equality of formulas only up to renaming of bound variables and often implicitly assumes that a bound variable is bound only once and does not occur free. Thus, if a free variable y of the term t is bound by quantifiers in B, an implicit renaming of the bound occurrences of y in B is admitted to enable backward application of the inference rule.

Generalized Rule of δ-Quantification: $\frac{A[B]}{A[\delta x.\, B]}$, where the variable x must not occur in the context $A[\ldots]$, and δ stands for $\forall$ if $[\ldots]$ denotes a positive position in $A[\ldots]$, and δ stands for $\exists$ if this position is negative. Moreover, both for soundness and for the reason mentioned above, we require that $[\ldots]$ does not occur in the scope of any quantifier in $A[\ldots]$.

Again, if x occurs in the context $A[\ldots]$, an implicit renaming of the bound occurrences of x in $\delta x.\, B$ is admitted to enable backward application.

Generalized Rule of Simplification: $\frac{A[B \circ B]}{A[B]}$, where $\circ$ stands for $\vee$ if $[\ldots]$ denotes a positive position in $A[\ldots]$, and $\circ$ stands for $\wedge$ if this position is negative.

To enable a forward application of the inference rule, the bound variables may be renamed such that the two occurrences of B become equal.

Moreover, the *Generalized Rule of γ-Simplification* is the sub-rule for the case that B is of the form $\exists y.C$ if $[\ldots]$ denotes a positive position in $A[\ldots]$, and of the form $\forall y.C$ if this position is negative.

To avoid the complication of quantifiers within a formula, where it is hard to keep track of the scope of each individual quantification, all quantifiers can be moved to the front, provided some caution is taken with the renaming of quantified variables. This is called the *prenex form* of a formula. The *anti-prenex form* results from the opposite transformation, i.e. from moving the quantifiers inwards as much as possible. Herbrand achieves these transformations with his *Rules of Passage*.

Rules of Passage: The following six logical equivalences may be used for rewriting from left to right (*prenex direction*) and from right to left (*anti-prenex direction*), resulting in twelve deep inference rules:

$$
\begin{array}{lrcl}
(1) & \neg\forall x.A & \Leftrightarrow & \exists x.\neg A \\
(2) & \neg\exists x.A & \Leftrightarrow & \forall x.\neg A \\
(3) & (\forall x.A) \vee B & \Leftrightarrow & \forall x.\,(A \vee B) \\
(4) & B \vee \forall x.A & \Leftrightarrow & \forall x.\,(B \vee A) \\
(5) & (\exists x.A) \vee B & \Leftrightarrow & \exists x.\,(A \vee B) \\
(6) & B \vee \exists x.A & \Leftrightarrow & \exists x.\,(B \vee A)
\end{array}
$$

Here, B is a formula in which the variable x does not occur. As explained above, if x occurs free in B, an implicit renaming of the bound occurrences of x in A is admitted to enable rewriting in prenex direction.

If we restrict the "Generalized" rules to outermost applications only (i.e. restrict A to be the empty context), we obtain the rules without the attribute "Generalized", i.e. the Rules of γ- and δ-Quantification and the Rule of Simplification.[62]

3.8 Normal Identities, Properties A, B, and C, and Herbrand Disjunction and Complexity

Key notions of Herbrand's thesis are *normal identity*, *Property A*, *Property B*, and *Property C*. Property C is the most important and the only one we need in this account.[63]

Herbrand's Property C was implicitly used already in [Löwenheim, 1915] and [Skolem, 1928], but as an explicit notion, it was first formulated in Herbrand's thesis. It is the main property of Herbrand's work and may well be called the central property of first-order logic, for reasons to be explained in the following.

In essence, Property C captures the following intuition taken from [Löwenheim, 1915]: Assuming the axiom of choice, the validity of a formula A is equivalent to the validity of its Skolemized form F. Moreover, the validity of F would be equivalent to the validity of the Herbrand expansion $F^{\mathcal{U}}$ for a universe $\mathcal{U}$, provided only that this expansion were a finite formula and did not vary over different universes. To provide this, we replace the semantical objects of the universe $\mathcal{U}$ with syntactical objects, namely the countable set of all terms, used as "place holders" or names. To get a *finite* formula, we again replace this set of terms, which is infinite in general, with the champ fini $\mathcal{T}_n$, as defined in § 3.5. If we can show $F^{\mathcal{T}_n}$ to be a

[62] The *Generalized* Rules of Quantification are introduced (under varying names) in [Heijenoort, 1968; Heijenoort, 1975; Heijenoort, 1982; Heijenoort, 1986c] and under the names (μ^*) and (ν^*) in [Hilbert and Bernays, 1968/70, Vol. II, p.166], but only in the second edition of 1970, not in the first edition of 1939. Herbrand had only the non-generalized versions of the Rules of Quantification and named them "First and Second Rule of Generalization", translated according to [Herbrand, 1971]. Note that the restrictions of the Generalized Rules of Quantification guarantee the equivalence of the generalized and the non-generalized versions by the Rules of Passage; cf. [Heijenoort, 1968, p. 6]. Herbrand's name for *modus ponens* is "Rule of Implication". Moreover, "(Generalized) Rule of Simplification" and "Rules of Passage" are Herbrand's names. All other names introduced in § 3.7 are our own invention to simplify our following presentation.

[63] The following are the definitions for the omitted notions *normal identity*, *Property A*, and *Property B* for a formula D. D is a *normal identity* if D has a linear proof starting with a sentential tautology, possibly followed by applications of the Rules of Quantification, and finally possibly followed by applications of the Rules of Passage. D has *Property A* if D has a linear proof starting with a sentential tautology, possibly followed by applications of the Rules of Quantification, and finally possibly followed by applications of the Rules of Passage and the Generalized Rule of Simplification. Herbrand's original definition of Property A is technically more complicated, but extensionally defines the same property and is also intensionally very similar. Finally, D has *Property B of order n* if E has Property C of order n, where E results from possibly repeated application of the Rules of Passage to D, in anti-prenex direction as long as possible.

sentential tautology for some positive natural number n, then we know that the γ-quantifications in F have solutions in any structure, so that F and A are valid.[64] Otherwise, the Löwenheim–Skolem Theorem says that A is invalid.

DEFINITION 3 Property C, Herbrand Disjunction, Herbrand Complexity.
Let A be a first-order formula, in which, without loss of generality, any bound variable is bound exactly once and does not occur again free, neither as a variable nor as a function symbol. Let F be the outer Skolemized form of A. Let n be a positive natural number. Let the champ fini $\mathcal{T}_n$ be formed over the function and free variable symbols occurring in F.
A has Property C of order n if the Herbrand expansion $F^{\mathcal{T}_n}$ is a sentential tautology.

The Herbrand expansion $F^{\mathcal{T}_n}$ is sententially equivalent to the so-called *Herbrand disjunction of A of order n*, which is the finite disjunction $\bigvee_{\sigma:\mathrm{Y}\to\mathcal{T}_n} E\sigma$, where Y is the set of bound (γ-) variables of F, and E results from F by removing all (γ-) quantifiers.

This form of representation can be used to define the *Herbrand complexity* of A, which is the minimal number of instances of E whose disjunction is a sentential tautology.[65] □

EXAMPLE 4 Property C, Herbrand Disjunction, Herbrand Complexity.
Let A be the following formula, which says that if we have transitivity and an upper bound of two elements, then we also have an upper bound of three elements:

$$\begin{array}{lll} & \forall a,b,c.\ (a \prec b \wedge b \prec c \Rightarrow a \prec c) & \\ \wedge & \forall x,y.\ \exists m.\ (x \prec m \wedge y \prec m) & (A) \\ \Rightarrow & \forall u,v,w.\ \exists n.\ (u \prec n \wedge v \prec n \wedge w \prec n) & \end{array}$$

The outer Skolemized form F of A is

$$\begin{array}{lll} & \forall a,b,c.\ (a \prec b \wedge b \prec c \Rightarrow a \prec c) & \\ \wedge & \forall x,y.\ (x \prec m^\delta(x,y) \wedge y \prec m^\delta(x,y)) & (F) \\ \Rightarrow & \exists n.\ (u^\delta \prec n \wedge v^\delta \prec n \wedge w^\delta \prec n) & \end{array}$$

The result of removing the quantifiers from F is the formula E:

$$\begin{array}{lll} & (a \prec b \wedge b \prec c \Rightarrow a \prec c) & \\ \wedge & x \prec m^\delta(x,y) \wedge y \prec m^\delta(x,y) & (E) \\ \Rightarrow & u^\delta \prec n \wedge v^\delta \prec n \wedge w^\delta \prec n & \end{array}$$

By semantical considerations it is obvious that a solution for n is $m^\delta(u^\delta, m^\delta(v^\delta, w^\delta))$. This is a term of height 3, which suggests that A has Property C of order 4. Let us show that this is indeed the case and that the Herbrand complexity of A is 2. To wit: Consider the following 2 substitutions:

---

[64] We have $F^{\mathcal{T}_n} \vdash A$, cf. Theorem 4 in [Heijenoort, 1975], which roughly is our Lemma 6.

[65] Herbrand has no name for *Herbrand disjunction* and does not use the notion of *Herbrand complexity*, which, however, is closely related to Herbrand's Fundamental Theorem, which says that the Herbrand complexity of A is always defined as a positive natural number, provided that $\vdash A$ holds. More formally, the *Herbrand complexity of A* is defined as the minimal cardinality $|S|$ such that, for some positive natural number m and some $S \subseteq \mathrm{Y} \to \mathcal{T}_m$, the finite disjunction $\bigvee_{\sigma\in S} E\sigma$ is a sentential tautology. It is useful in the comparison of logical calculi w.r.t. their smallest proofs for certain generic sets of formulas, cf. e.g. [Baaz and Fermüller, 1995].

$$\begin{array}{lllll} \{ & a\mapsto v^\delta, & b\mapsto m^\delta(v^\delta,w^\delta), & c\mapsto m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)), & \\ & x\mapsto v^\delta, & y\mapsto w^\delta, & n\mapsto m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)) & \}; \\ \{ & a\mapsto w^\delta, & b\mapsto m^\delta(v^\delta,w^\delta), & c\mapsto m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)), & \\ & x\mapsto u^\delta, & y\mapsto m^\delta(v^\delta,w^\delta), & n\mapsto m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)) & \}. \end{array}$$

Indeed, if we normalize the Herbrand disjunction generated by these two substitutions to a disjunctive normal form (i.e. a disjunctive set of conjunctions) we get the following sentential tautology.

$$\begin{array}{l} \{v^\delta \prec m^\delta(v^\delta,w^\delta) \wedge m^\delta(v^\delta,w^\delta) \prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)) \wedge v^\delta \not\prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)), \\ \ w^\delta \prec m^\delta(v^\delta,w^\delta) \wedge m^\delta(v^\delta,w^\delta) \prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)) \wedge w^\delta \not\prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)), \\ \ v^\delta \not\prec m^\delta(v^\delta,w^\delta),\ \ w^\delta \not\prec m^\delta(v^\delta,w^\delta), \\ \ u^\delta \not\prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)),\ \ m^\delta(v^\delta,w^\delta) \not\prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)), \\ \ u^\delta \prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)) \wedge v^\delta \prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta)) \wedge w^\delta \prec m^\delta(u^\delta,m^\delta(v^\delta,w^\delta))\} \end{array} \quad (C) \quad \square$$

The different treatment of δ-quantifiers and γ-quantifiers in Property C, namely by Skolemization and Herbrand expansion, respectively, as found in [Skolem, 1928] and [Herbrand, 1930], rendered the reduction to sentential logic by hand (or actually today, on a computer) practically executable for the first time.[66] This different treatment of the two kinds of quantification is inherited from the Peirce–Schröder tradition[40] which came on Herbrand via Löwenheim and Skolem. Russell and Hilbert had already merged that tradition with the one of Frege, sometimes emphasizing their Frege heritage over one of Peirce and Schröder.[67] It was Herbrand who completed the bridge between these two traditions with his Fundamental Theorem, as depicted in § 3.10 below.

3.9 Herbrand's False Lemma

For a given positive natural number n, *Herbrand's (False) Lemma* says that Property C of order n is invariant under the application of the Rules of Passage.

The basic function of Herbrand's False Lemma in the proof of Herbrand's Fundamental Theorem is to establish the logical equivalence of Property C of a formula A with Property C of the *prenex* and *anti-prenex forms* of A, cf. § 3.7.

Herbrand's Lemma is wrong because the Rules of Passage may change the outer Skolemized form. This happens whenever a γ-quantifier binding x is moved over a binary operator whose unchanged operand B contains a δ-quantifier.[68]

To find a counterexample for Herbrand's Lemma for the case of Property C of order 2, let us consider moving out the γ-quantifier "$\exists x.$" in the valid formula

[66] For instance, the elimination of both γ- and δ-quantifiers with the help of Hilbert's ε-operator suffers from an exponential complexity in formula size. As a result, already small formulas grow so large that the mere size makes them inaccessible to human inspection; and this is still the case for the term-sharing representation of ε-terms of [Wirth, 2008a].

[67] While this emphasis on Frege will be understood by everybody who ever had the fascinating experience of reading Frege, it put some unjustified bias to the historiography of modern logic, still present in the selection of the famous source book [Heijenoort, 1971a], cf. e.g. [Anellis, 1992, Chapter 3].

[68] Here we use the same meta variables as in our description of the Rules of Passage in § 3.7 and assume that x does not occur free in B.

$$(\exists x.\,\mathsf{P}(x)) \vee \neg\exists y.\,\mathsf{P}(y).$$

The (outer) Skolemized form of this formula is

$$(\exists x.\,\mathsf{P}(x)) \vee \neg\mathsf{P}(y^\delta).$$

The Herbrand disjunction over the single substitution $\{x\mapsto y^\delta\}$ is a sentential tautology. The *outer* Skolemized form after moving out the "$\exists x.$" is

$$\exists x.\ \big(\ \mathsf{P}(x) \vee \neg\mathsf{P}(y^\delta(x))\ \big).$$

To get a sentential tautology again, we now have to take the Herbrand disjunction over both $\{x\mapsto y^\delta(l)\}$ and $\{x\mapsto l\}$ (instead of the single $\{x\mapsto y^\delta\}$), for the lexicon l.

This, however, is not really a counterexample for Herbrand's Lemma because Herbrand treated the lexicon l as a variable and defined the height of a Skolem constant to be 1, and the height of a variable to be 0, so that $|y^\delta| = 1 = |y^\delta(l)|$. As free variables and Skolem constants play exactly the same rôle, this definition of height is a bit unintuitive and was possibly introduced to avoid this counterexample. But now for the similar formula

$$\big(\ (\exists x.\,\mathsf{P}(x)) \wedge \forall y.\,\mathsf{Q}(y)\ \big) \ \vee\ \neg(\exists x.\,\mathsf{P}(x)) \ \vee\ \neg\forall y.\,\mathsf{Q}(y)$$

after moving the γ-quantifier "$\exists x.$" out, we have to apply

$$\{\ x \mapsto x^\delta,\quad y \mapsto y^\delta(x^\delta)\ \}$$

instead of

$$\{\ x \mapsto x^\delta,\quad y \mapsto y^\delta\ \}$$

to get a sentential tautology, and we have $|y^\delta| = 1$ and $|y^\delta(x^\delta)| = 2$, and thus $y^\delta \in \mathcal{T}_2$, but $y^\delta(x^\delta) \notin \mathcal{T}_2$. This means that Property C of order 2 varies under a single application of a Rule of Passage, and thus we have a proper counterexample for Herbrand's False Lemma here.

In 1939, Bernays remarked that Herbrand's proof is hard to follow[69] and — for the first time — published a sound proof of a version of Herbrand's Fundamental Theorem which is restricted to prenex form, but more efficient in the number of terms that have to be considered in a Herbrand disjunction than Herbrand's quite global limitation to *all terms t with* $|t| < n$, related to Property C of order n.[70]

According to a conversation with Heijenoort in autumn 1963,[71] Gödel noticed the lacuna in the proof of Herbrand's False Lemma in 1943 and wrote a private note, but did not publish it. While Gödel's documented attempts to construct a counterexample to Herbrand's False Lemma failed, he had actually worked out a *correction* of Herbrand's False Lemma, which is sufficient for the proof of Herbrand's Fundamental Theorem.[72]

In 1962, when Gödel's correction was still unknown, a young student, Peter B. Andrews, had the audacity to tell his advisor Alonzo Church (1903–1995) that there seemed to be a gap in the proof of Herbrand's False Lemma. Church sent Andrews to Burton Dreben (1927–1999), who finally came up with a counterexample. And then Andrews constructed a simpler counterexample (essentially the one we presented above) and joint work found a correction similar to Gödel's,[73] to which

[69] In [Hilbert and Bernays, 1968/70, Vol. II], in Note 1 on p.158 of the 1939 edition (p.161 of the 1970 edition), we read: "Die Herbrandsche Beweisführung ist schwer zu verfolgen"

[70] Cf. § 3.3 of the 1939 edition of [Hilbert and Bernays, 1968/70, Vol. II]. In the 1970 edition, Bernays also indicates how to remove the restriction to prenex formulas.

[71] Cf. [Herbrand, 1968, p. 8, Note j].

[72] Cf. [Goldfarb, 1993].

we will refer to as *Gödel's and Dreben's correction* in § 3.11.

Roughly speaking, the corrected lemma says that — to keep Property C of A invariant under (a single application of) a Rule of Passage — we may have to step from order n to order $n\,(N^r{+}1)^n$. Here r is the number of γ-quantifiers in whose scope the Rule of Passage is applied and N is the cardinality of $\mathcal{T}_n$ for the function symbols in the outer Skolemized form of A.[74] This correction is not particularly elegant because — iterated several times until a prenex form is reached — it can lead to pretty high orders. Thus, although this correction serves well for soundness and finitism, it results in a complexity that is unacceptable in practice (e.g. in automated reasoning) already for small non-prenex formulas.

The problems with Herbrand's False Lemma in the step from Property C to a proof without *modus ponens* in his Fundamental Theorem (cf. § 3.11) result primarily[75] from a detour over prenex form, which was standard at Herbrand's time. Löwenheim and Skolem had always reduced their problems to prenex forms of various kinds. The reduction of a proof task to prenex form has several disadvantages, however, such as serious negative effects on proof complexity.[76] If Jacques Herbrand had known of his flaw, he would probably have avoided the whole detour over prenex forms here, namely in the form of what we will call *Heijenoort's correction*, which avoids an intractable and unintuitive[77] rise in complexity, cf. § 3.11.[78]

[73] Cf. [Andrews, 2003], [Dreben *et al.*, 1963], [Dreben and Denton, 1963].

[74] Cf. [Dreben and Denton, 1963, p. 393].

[75] Secondarily, the flaw in Herbrand's False Lemma is a peculiarity of the *outer* Skolemized form. For the *inner* Skolemized form (cf. Note 59), moving γ-quantifiers with the Rules of Passage cannot change the number of arguments of the Skolem functions. This does not help, however, because, for the inner Skolemized form, moving a δ-quantifier may change the number of arguments of its Skolem function if the Rule of Passage is applied within the scope of a γ-quantifier whose bound variable occurs in B but not in A.[68] The inner Skolemized form of $\exists y_1.\ \forall z_1.\ \mathsf{Q}(y_1, z_1) \vee \exists y_2.\ \forall z_2.\ \mathsf{Q}(y_2, z_2)$ is $\exists y_1.\ \mathsf{Q}(y_1, z_1^\delta(y_1)) \vee \exists y_2.\ \mathsf{Q}(y_2, z_2^\delta(y_2))$, but the inner Skolemized form of any prenex form has a *binary* Skolem function, unless we use Henkin quantifiers as found in Hintikka's first-order logic, cf. [Hintikka, 1966].

[76] Cf. e.g. [Baaz and Fermüller, 1995]; [Baaz and Leitsch, 1995].

[77] Unintuitive e.g. in the sense of [Tait, 2006].

[78] The first published hints on *Heijenoort's correction* are [Heijenoort, 1971a, Note 77, p. 555] and [Herbrand, 1971, Note 60, p.171]. On page 99 of [Heijenoort, 1986c], without giving a definition, Heijenoort speaks of generalized versions (which Herbrand did not have) of the rules of "existentialization and universalization", which we have formalized in our Generalized Rules of Quantification in § 3.7. Having studied Herbrand's Ph.D. thesis [Herbrand, 1930] and [Heijenoort, 1968; Heijenoort, 1975], what Heijenoort's generalized rules must look like can be inferred from the following two facts: Herbrand has a generalized version of his Rule of Simplification in addition to a non-generalized one. Rewriting with the Generalized Rule of γ-Quantification within the scope of quantifiers would not permit Herbrand's constructive proof of his Fundamental Theorem.

Note that Heijenoort's correction avoids the detour over the Extended First ε-Theorem of the proof of Bernays mentioned above. Moreover, Heijenoort gets along without Herbrand's complicated prenex forms with raised γ-multiplicity, which are required for Herbrand's definition of Property A.

Notice, however, that Gödel's and Dreben's correction is still needed for the step from a proof with *modus ponens* to Property C, i.e. from Statement 4 to Statement 1 in Theorem 5. As the example on top of page 201 in [Herbrand, 1971] shows, an intractable increase of the order of Property C cannot be avoided in general for an inference step by *modus ponens*.

3.10 The Fundamental Theorem

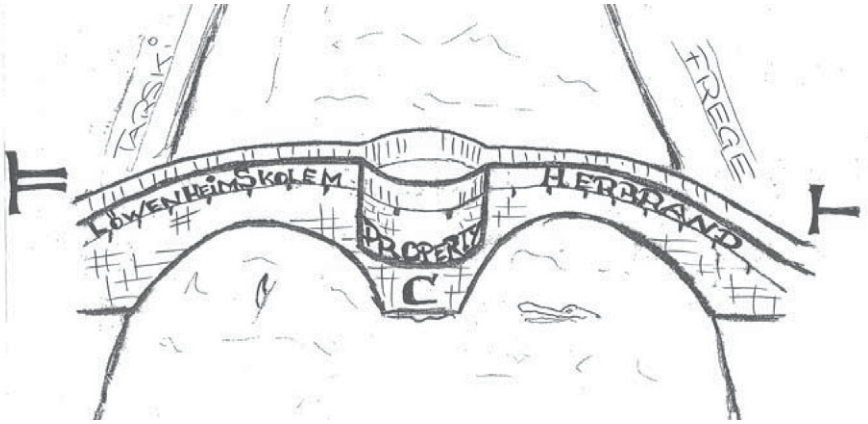
The Fundamental Theorem of Jacques Herbrand is not easy to comprehend at first, because of its technical nature, but it rests upon a basic intuitive idea, which turned out to be one of the most profound insights in the history of logic.

We know — and so did Herbrand — that sentential logic is decidable: for any given sentential formula, we could, for instance, use truth-tables to decide its validity. But what about a first-order formula with quantifiers?

There is Löwenheim's and Skolem's observation that $\forall x.\, \mathsf{P}(x)$ in the context of the existentially quantified variables $y_1, \ldots, y_n$ stands for $\mathsf{P}(x^\delta(y_1, \ldots, y_n))$ for an arbitrary Skolem function $x^\delta(\cdots)$, as outlined in § 3.6. This gives us a formula with existential quantifiers only. Now, taking the Herbrand disjunction, an existentially quantified formula can be shown to be valid, if we find a finite set of names denoting elements from the domain to be substituted for the existentially quantified variables, such that the resulting sentential formula is truth-functionally valid. Thus, we have a model-theoretic argumentation how to reduce a given first-order formula to a sentential one. The semantical elaboration of this idea is due to Löwenheim and Skolem, and this was known to Herbrand.

But what about an *actual proof* within a given calculus for a given formula? The affirmative answer to this question is the essence of Herbrand's Fundamental Theorem and the technical device, by which we can eliminate a switch of quantifiers (such as $\exists y.\, \forall x.\, \mathsf{Q}(x, y)$ of § 3.6) is captured in his Property C.

Thus, if we want to cross the river that divides the land of *valid* first-order formulas from the land of *provable* ones, it is the sentential Property C that stands firm in the middle of that river and holds the bridge, whose first half was built by Löwenheim and Skolem and the other by Herbrand:



Herbrand's Fundamental Theorem shows that if a formula A has Property C of some order n — i.e., by the Löwenheim–Skolem Theorem, if A is a valid ($\models A$) — then we not only *know of the existence* of a proof in any of the standard proof calculi ($\vdash A$), but we can actually *construct* a proof for A in Herbrand's calculus from a given n. The proof construction process is guided by the champ fini of order n, whose size determines the multiplicities of γ-quantifiers and whose elements are the terms substituted as witnesses in the γ-Quantification steps. That proof begins with a sentential tautology and may use the Rules of γ- and δ-Quantification, the Generalized Rule of Simplification, and the Rules of Passage.

Contrary to what Herbrand's False Lemma implies, a detour over a prenex form of A dramatically increases the order of Property C and thus the length of that proof, cf. § 3.9. Heijenoort, however, observed that this rise of proof length can be overcome by avoiding the problematic Rules of Passage with the help of deep (or Generalized) quantification rules, which may introduce quantifiers deep within formulas ("Heijenoort's correction"). We have included these considerations into our statement of Herbrand's Fundamental Theorem.

THEOREM 5 Herbrand's Fundamental Theorem.
Let A be a first-order formula in which each bound variable is bound by a single quantifier and does not occur free. The following five statements are logically equivalent. Moreover, we can construct a witness for any statement from a witness of any other statement.

1. *A has Property C of order n for some positive natural number n.*
2. *We can derive A from a sentential tautology, starting possibly with applications of the Generalized Rules of γ- and δ-Quantification, which are then possibly followed by applications of the Generalized Rule of γ-Simplification.*
3. *We can derive A from a sentential tautology, starting possibly with applications of the Rules of γ- and δ-Quantification, which are then possibly followed by applications of the Generalized Rule of γ-Simplification and the Rules of Passage.*
4. *We can derive A from a sentential tautology with the Rules of γ- and δ-Quantification, the Rule of Simplification, the Rules of Passage, and Modus Ponens.*
5. *We can derive A in one of the standard first-order calculi of Principia Mathematica or of the Hilbert school.*[79] □

The following deserves emphasis: The derivations in the above Statements 2 to 5 as well as the number n of Statement 1 can be *constructed* from each other; and this construction is finitistic in the spirit of Herbrand's basic beliefs in the nature of proof theory and meta-mathematics. Statement 3 and Herbrand's Property A are extensionally equal and intensionally very close to each other. Statement 2 is due to Heijenoort's correction, cf. §§ 3.9 and 3.11.

[79] Cf. [Whitehead and Russell, 1910–1913, *10], [Hilbert and Bernays, 1968/70, Vol. II, Supplement I D], respectively.

3.11 Modus Ponens Elimination

The following lemma provides the step from Statement 1 to Statement 2 of Theorem 5 with additional details exhibiting an elimination of *modus ponens* similar to the Cut elimination in Gentzen's Hauptsatz.

We present the lemma in parallel both in the version of Gödel's and Dreben's correction and of Heijenoort's correction.[80] To melt these two versions into one, we underline the parts that are just part of Heijenoort's correction and overline the parts that result from Gödel's and Dreben's. Thus, the lemma stays valid if we omit either the underlined or else the overlined part of it, but not both.

LEMMA 6 *Modus Ponens* Elimination.

Let A be a first-order formula $\overline{\text{in prenex form}}$ *in which each bound variable is bound by a single quantifier and does not occur free. Let F be the outer Skolemized form of A. Let Y be the set of bound (γ-) variables of F. Let E result from F by removing all (γ-) quantifiers. Let n be a positive natural number. Let the champ fini $\mathcal{T}_n$ be formed over the function and free variable symbols occurring in F.*
If A has Property C of order n, then we can construct a derivation of A of the following form, in which we read any term starting with a Skolem function as an atomic variable:

Step 1: *We start with* $\underline{\text{a sentential tautology whose disjunctive normal form is a re-ordering of a disjunctive normal form of}}$ *the sentential tautology* $\bigvee_{\sigma:\mathrm{Y}\to\mathcal{T}_n} E\sigma$.

Step 2: *Then we may repeatedly apply the* $\underline{\text{Generalized}}$ *Rules of γ- and δ-Quantification.*

Step 3: *Then, (after renaming all bound δ-variables) we may repeatedly apply the Generalized Rule of γ-Simplification.* □

Obviously, there is no use of *modus ponens* in such a proof, and thus, it is linear, i.e. written as a tree, it has no branching. Moreover, all function and predicate symbols within this proof occur already in A, and all formulas in the proof are similar to A in the sense that they have the so-called *"sub"-formula property.*

EXAMPLE 7 *Modus Ponens* Elimination. *(continuing Example 4)*
Let us derive the formula A of Example 4 in § 3.8. As A is not in prenex form we have to apply the version of Lemma 6 without the overlined part. As explained in Example 4, A has Property C of order n for $n=4$, and the result of removing the quantifiers from the outer Skolemized form of A is the formula E:

$$\begin{array}{ll} & (a \prec b \wedge b \prec c \Rightarrow a \prec c) \\ \wedge & x \prec m^{\delta}(x,y) \wedge y \prec m^{\delta}(x,y) \\ \Rightarrow & u^{\delta} \prec n \wedge v^{\delta} \prec n \wedge w^{\delta} \prec n \end{array} \qquad (E)$$

[80] Cf. § 3.9. We present Heijenoort's correction actually in form of Theorem 4 in [Heijenoort, 1975] with a slight change, which becomes necessary for our use of Herbrand disjunction instead of the Herbrand expansion, namely the addition of the underlined part of Step 1 in Lemma 6.

Let N denote the cardinality of $\mathcal{T}_n$. Let $\mathcal{T}_n = \{t_1, \ldots, t_N\}$.

For the case of $n=4$, we have $N = 3 + 3^2 + (3+3^2)^2 = 156$, and, for $\mathrm{Y} := \{a, b, c, n, x, y\}$, the Herbrand disjunction $\bigvee_{\sigma:\mathrm{Y}\to\mathcal{T}_4} E\sigma$ has $N^{|Y|}$ elements, i.e. more than 10^{13}. Thus, we had better try a reduction proof here,[81] applying the inference rules backwards, and be content with arriving at a sentential tautology which is a sub-disjunction of a re-ordering of a disjunctive normal form of $\bigvee_{\sigma:\mathrm{Y}\to\mathcal{T}_4} E\sigma$.

As the backwards application of the Generalized Rule of γ-Quantification admits only a single (i.e. linear) application of each γ-quantifier (or each "lemma"), and as we will have to apply both the first and the second line of A twice, we first increase the γ-multiplicity of the top γ-quantifiers of these two lines to two. This is achieved by applying the Generalized Rule of γ-Simplification twice backwards to A, resulting in:[82]

$$\begin{array}{ll} & \forall a,b,c.\ (a \prec b \wedge b \prec c \Rightarrow a \prec c) \\ \wedge & \forall a,b,c.\ (a \prec b \wedge b \prec c \Rightarrow a \prec c) \\ \wedge & \forall x,y.\ \exists m.\ (x \prec m \wedge y \prec m) \\ \wedge & \forall x,y.\ \exists m.\ (x \prec m \wedge y \prec m) \\ \Rightarrow & \forall u,v,w.\ \exists n.\ (u \prec n \wedge v \prec n \wedge w \prec n) \end{array}$$

Renaming the bound δ-variables to some terms from $\mathcal{T}_n$, and applying the Generalized Rule of δ-Quantification three times backwards in the last line, we get:

$$\begin{array}{ll} & \forall a,b,c.\ (a \prec b \wedge b \prec c \Rightarrow a \prec c) \\ \wedge & \forall a,b,c.\ (a \prec b \wedge b \prec c \Rightarrow a \prec c) \\ \wedge & \forall x,y.\ \exists \boxed{m^\delta(v^\delta, w^\delta)}.\ (x \prec \boxed{m^\delta(v^\delta, w^\delta)} \wedge y \prec \boxed{m^\delta(v^\delta, w^\delta)}) \\ \wedge & \forall x,y.\ \exists \boxed{m^\delta(u^\delta, m^\delta(v^\delta, w^\delta))}.\ \left(\begin{array}{ll} & x \prec \boxed{m^\delta(u^\delta, m^\delta(v^\delta, w^\delta))} \\ \wedge & y \prec \boxed{m^\delta(u^\delta, m^\delta(v^\delta, w^\delta))} \end{array} \right) \\ \Rightarrow & \exists n.\ (\boxed{u^\delta} \prec n \wedge \boxed{v^\delta} \prec n \wedge \boxed{w^\delta} \prec n) \end{array}$$

The boxes indicate that the enclosed term actually denotes an atomic variable whose structure cannot be changed by a substitution. By this nice trick of taking outermost Skolem terms as names for variables, Herbrand avoids the hard task of giving semantics to Skolem functions, cf. § 3.12.[83]

[81] As Herbrand's proof [Herbrand, 1971, p.170] proceeds reductively too, we explain the proof in parallel to the development of our example, in footnotes. This proof is interesting by itself and similar to the later proof of the Second ε-Theorem in § 3.1 of [Hilbert and Bernays, 1968/70, Vol. II].

[82] To arrive at the full Herbrand disjunction $\bigvee_{\sigma:\mathrm{Y}\to\mathcal{T}_n} E\sigma$, Herbrand's proof requires us to apply the Rule of Simplification top-down at any occurrence of a γ-quantifier N times, and the idea is to substitute t_i for the i^{th} occurrence of this γ-quantifier on any branch.

[83] According to Herbrand's proof we would have to replace any bound δ-variable x with its Skolem term $x^\delta(t_{i_0}, \ldots, t_{i_k})$, provided that $i_0, \ldots, i_k$ denotes the branch on which this δ-quantifier occurs w.r.t. the previous step of raising each γ-multiplicity to N, described in Note 82.

We apply the Generalized Rule of γ-Quantification four times backwards, resulting in application of

$$\{x \mapsto \boxed{v^\delta},\ \ y \mapsto \boxed{w^\delta}\}$$

to the third line and

$$\{x \mapsto \boxed{u^\delta},\ \ y \mapsto \boxed{m^\delta(v^\delta, w^\delta)}\}$$

to the fourth line. This yields:[84]

$$\begin{array}{lll}
 & \forall a,b,c.\ (a \prec b \wedge b \prec c \Rightarrow a \prec c) & \\
\wedge & \forall a,b,c.\ (a \prec b \wedge b \prec c \Rightarrow a \prec c) & \\
\wedge & \exists \boxed{m^\delta(v^\delta, w^\delta)}.\ (\boxed{v^\delta} \prec \boxed{m^\delta(v^\delta, w^\delta)} \wedge \boxed{w^\delta} \prec \boxed{m^\delta(v^\delta, w^\delta)}) & \\
\wedge & \exists \boxed{m^\delta(u^\delta, m^\delta(v^\delta, w^\delta))}. & \left(\begin{array}{ll} & \boxed{u^\delta} \prec \boxed{m^\delta(u^\delta, m^\delta(v^\delta, w^\delta))} \\ \wedge & \boxed{m^\delta(v^\delta, w^\delta)} \prec \boxed{m^\delta(u^\delta, m^\delta(v^\delta, w^\delta))} \end{array} \right) \\
\Rightarrow & \exists n.\ (\boxed{u^\delta} \prec n \wedge \boxed{v^\delta} \prec n \wedge \boxed{w^\delta} \prec n) &
\end{array}$$

Applying (always backwards) the Generalized Rule of δ-Quantification twice and the Generalized Rule of γ-Quantification seven times, and then dropping the boxes (as they are irrelevant for sentential reasoning without substitution) and rewriting it all into a disjunctive list of conjunctions, we arrive at the disjunctive set C of Example 4, which is a sentential tautology. Moreover, as a list, C is obviously a re-ordered sublist of a disjunctive normal form of $\bigvee_{\sigma:\mathrm{Y}\to\mathcal{T}_4} E\sigma$. □

In the time before Herbrand's Fundamental Theorem, a calculus was basically a means to describe a set of theorems in a semi-decidable and theoretical fashion. In Hilbert's calculi, for instance, the search for concrete proofs is very hard. Contrary to most other Hilbert-style calculi, the normal form of proofs given in Statement 2 of Theorem 5, however, supports the search for reductive proofs: Methods of human[85] and automatic[86] proof search may help us to find simple proofs in this

[84] Note that the terms to be substituted for a bound γ-variable, say y, in such a reduction step can always be read out from any bound δ-variable in its scope: If there are j γ-quantifiers between the quantifier for y inclusively and the quantifier for the δ-variable, the value for y is the j^{th} argument of the bound δ-variable, counting from the last argument backwards.
For instance, in the previous reduction step, the variable y in the third line was replaced with $\boxed{w^\delta}$, the last argument of the bound δ-variable $\boxed{m^\delta(v^\delta, w^\delta)}$, being first in the scope of y.
This property is obvious from Herbrand's proof but hard to express as a property of proof normalization.
Moreover, this property is useful in Herbrand's proof for showing that the side condition of the Rule of γ-Quantification $\frac{B\{x\mapsto t\}}{\exists x.\ B}$ is always satisfied, even for a certain prenex form. Indeed, γ-variables never occur in the replacement t and the height of t is strictly smaller than the height of all bound δ-variables in the scope B, so that no free variable in t can be bound by quantifiers in B; cf. § 3.7.

[85] Roughly speaking, we may do a proof by hand, count the lemma applications and remember their instantiations, and then try to construct a formal normal form proof accordingly, just as we have done in Example 7. See [Wirth, 2004] for more on this.

[86] Roughly speaking, we may compute the connections and search for a reductive proof in the style of say [Wallen, 1990], which we then transform into a proof in the normal form of Statement 2 of Theorem 5.

normal form. Lemma 6 gives us the means to search successfully for simple proofs in a formal calculus by hand (or actually today, on a computer), just as we have done in Example 7.[87]

The normal form of proofs — as given by Lemma 6 — eliminates detours via *modus ponens* in a similar fashion as Gentzen's Hauptsatz eliminates the Cut. It is remarkable not only because it establishes a connection between Skolem terms and free variables without using any semantics for Skolem functions (and thereby, without using the axiom of choice). It also seems to be the first time that a normal form of proofs is shown to exist in which *different phases* are considered. Even with Gentzen's Verschärfter Hauptsatz following Herbrand in this aspect some years later, the concrete form of Herbrand's normal form of proofs remains important to this day, especially in the form of Heijenoort's correction, cf. § 3.9. The manner in which modern sequent, tableau, and matrix calculi organize proof search[88] does not follow the Hilbert school and their ε-elimination theorems, but Gentzen's and Herbrand's calculi. Moreover, regarding their Skolemization, their deep inference,[89] and their focus on γ-quantifiers and their multiplicity, these modern proof-search calculi are even more in Herbrand's tradition than in Gentzen's.

3.12 The Löwenheim–Skolem Theorem and Herbrand's Finitistic Notion of Falsehood in an Infinite Domain

Let A be a first-order formula whose terms have a height not greater than m. Herbrand defines that A *is false in an infinite domain* if A does not have Property C of order p for any positive natural number p.

If, for a given positive natural number p, the formula A does not have Property C of order p, then we can construct a finite structure over the domain $\mathcal{T}_{p+m}$ which falsifies $A^{\mathcal{T}_p}$; cf. § 3.5. Thus, instead of requiring a single infinite structure in which $A^{\mathcal{T}_p}$ is false for any positive natural number p, Herbrand's notion of falsehood in an infinite domain only provides us, for each p, with a finite structure in which $A^{\mathcal{T}_p}$ is false. Herbrand explicitly points out that these structures do not have to be extensions of each other. From a given falsifying structure for some p one can, of course, generate falsifying structures for each $p' < p$ by restriction to $\mathcal{T}_{p'+m}$. Herbrand thinks, however, that to require an infinite sequence of structures to be a sequence of extensions would necessarily include some form of the axiom of choice, which he rejects out of principle. Moreover, he writes that the basic prerequisites of the Löwenheim–Skolem Theorem are generally misunderstood, but does not make this point clear.

[87] Even without the avoidance of the detour over prenex forms due to Heijenoort's correction, this already holds for the normal form given by Statement 3 of Theorem 5, which is extensionally equal to Herbrand's original Property A. The next further steps to improve this support for proof search would be sequents and *free* γ- and δ-variables; cf. e.g. [Wirth, 2004; 2008a].

[88] Cf. e.g. [Wallen, 1990], [Wirth, 2004], [Autexier, 2005].

[89] Note that although the deep inference rules of *Generalized* Quantification are an extension of Herbrand's calculi by Heijenoort, the deep inference rules of Passage and of Generalized Simplification are Herbrand's original contributions.

It seems that Herbrand reads [Löwenheim, 1915] as if it would be a paper on provability instead of validity, i.e. that Herbrand confuses Löwenheim's '$\models$' with his '$\vdash$'.

All in all, this is, on the one hand, so peculiar and, on the other hand, so relevant for Herbrand's finitistic views of logic and proof theory that some quotations may illuminate the controversy.

> "(On remarquera que cette définition diffère de la définition qu'on pourrait croire la plus naturelle seulement par le fait que, quand le nombre p augmente, le nouveau champ C' et les nouvelles valeurs ne peuvent pas forcément être considérés comme un 'prolongement' des anciens; mais cependant, la connaissance de C' et des valeurs pour un nombre p déterminé, entraîne celle d'un champ et de valeurs convenant pour tout nombre inférieur; seul donc un 'principe de choix' pourrait conduire à prendre un système de valeurs fixe dans un champ infini)."[90]

And after defining the dual notion for unsatisfiability instead of validity, Herbrand continues:

> "Il est absolument nécessaire de prendre de telles définitions, pour donner un sens précis aux mots: 'vrai dans un champ infini', qui ont souvent été employés sans explication suffisante, et pour justifier une proposition à laquelle on fait souvent allusion, démontrée par *Löwenheim*,[1] sans bien remarquer que cette proposition n'a aucun sens précis sans définition préalable, et que la démonstration de *Löwenheim* est au surplus totalement insuffisante pour notre but (voir 6.4)."[91]

[90] Cf. [Herbrand, 1930, p.109]. Without the comma after "déterminé" also in: [Herbrand, 1968, p.135f.]
"We observe that this definition differs from the definition that would seem the most natural only in that, as the number p increases, the new domain C' and the new values need not be regarded as forming an 'extension' of the previous ones. Clearly, if we know C' and the values for a given number p, then for each smaller number we know a domain and values that answer to the number; but only a 'principle of choice' could lead us to take a fixed system of values in an infinite domain." [Herbrand, 1971, p.165, translation by Dreben and Heijenoort]

[91] Cf. [Herbrand, 1930, p.110]. Without the signs after "définitions", "mots", and "préalable" and the emphasis on "*Löwenheim*", but with the correct "6.2" instead of the misprint "6.4" also in: [Herbrand, 1968, p.136f.]
"It is absolutely necessary to adopt such definitions if we want to give a precise sense to the words 'true in an infinite domain', words that have frequently been used without sufficient explanation, and also if we want to justify a proposition proved by Löwenheim, a proposition to which many refer without clearly seeing that Löwenheim's proof is totally inadequate for our purposes (see 6.4) and that, indeed, the proposition has no precise sense until such a definition has been given." [Herbrand, 1971, p.166, translated by Dreben and Heijenoort]

Herbrand's Fundamental Theorem equates *provability* with Property C, whereas the Löwenheim–Skolem Theorem equates *validity* with Property C. Thus, it is not the case that Herbrand somehow corrected Löwenheim. Instead, the Löwenheim–Skolem Theorem and Herbrand's Fundamental Theorem had better be looked upon as a bridge from validity to provability with two arcs and Property C as the eminent pillar in the middle of the river, offering a magnificent view from the bridge on properties of first-order logic; as depicted in § 3.10. And this was probably also Herbrand's view when he correctly wrote:

> "Nous pouvons dire que la démonstration de *Löwenheim* était suffisante en Mathématiques; mais il nous a fallu, dans ce travail, la rendre 'métamathématique' (voir l'introduction), pour qu'elle nous soit de quelque utilité."[92]

Moreover, Herbrand criticizes Löwenheim for not showing the consistency of first-order logic, but this, of course, was never Löwenheim's concern.

The mathematically substantial part of Herbrand's critique of Löwenheim refers to the use of the axiom of choice in Löwenheim's proof of the Löwenheim–Skolem Theorem.

The Löwenheim–Skolem Theorem as found in many textbooks, such as [Enderton, 1972, p.141], says that any satisfiable set of first-order formulas is satisfiable in a countable structure. In [Löwenheim, 1915], however, we only find a dual statement, namely that any invalid first-order formula has a denumerable countermodel. Moreover, what is actually proved, read charitably,[93] is the following stronger theorem:

THEOREM 8 Löwenheim–Skolem Theorem à la Löwenheim [1915].
Let us assume the axiom of choice. Let A be a first-order formula.

1. *If A has Property C of order p for some positive natural number p, then $\models A$.*
2. *If A does not have Property C of order p for any positive natural number p, then we can construct a sequence of partial structures $\mathcal{S}_i$ that converges to a structure $\mathcal{S}'$ with a denumerable universe such that $\not\models_{\mathcal{S}'} A$.* □

[92] Cf. [Herbrand, 1930, p.118]. Without the emphasis on "*Löwenheim*" and with "mathématiques" instead of "Mathématiques" and "l'Introduction)" instead of "l'introduction)," also in: [Herbrand, 1968, p.144].
"We could say that Löwenheim's proof was sufficient in mathematics; but, in the present work, we had to make it 'metamathematical' (see Introduction) so that it would be of some use to us."
[Herbrand, 1971, p.176, translated by Dreben and Heijenoort]

[93] As Löwenheim's paper lacks some minor details, there is an ongoing discussion whether its proof of the Löwenheim–Skolem Theorem is complete and what is actually shown. Our reading of the proof of the Löwenheim–Skolem Theorem in [Löwenheim, 1915] is a standard one. Only in [Skolem, 1941, p. 26ff.] and [Badesa, 2004, § 6.3.4] we found an incompatible reading, namely that — to construct $\mathcal{S}'$ of Item 2 of Theorem 8 — Löwenheim's proof requires an additional falsifying structure of arbitrary cardinality to be given in advance. The similarity of our presentation with Herbrand's Fundamental Theorem, however, is in accordance with [Skolem, 1941, p. 30], but not with [Badesa, 2004, p.145]. The relation of Herbrand's Fundamental Theorem to the Löwenheim–Skolem Theorem is further discussed in [Anellis, 1991]. Cf. also our Note 98.

As Property C of order p can be effectively tested for $p = 1, 2, 3, \ldots$, Löwenheim's proof provides us with a complete proof procedure which went unnoticed by Skolem as well as the Hilbert school. Indeed, there is no mention in the discussion of the completeness problem for first-order logic in [Hilbert and Ackermann, 1928, p. 68], where it is considered as an open problem.[94]

Thus, for validity instead of provability, Gödel's Completeness Theorem[95] is contained already in [Löwenheim, 1915]. Gödel has actually acknowledged this for the version of the proof of the Löwenheim–Skolem Theorem in [Skolem, 1923b].[96]

Note that the convergence of the structures $\mathcal{S}_i$ against $\mathcal{S}'$ in Theorem 8 is hypothetical in two aspects: First, as validity is not co-semi-decidable, in general we can never positively know that we are actually in Case 2 of Theorem 8, i.e. that a convergence toward $\mathcal{S}'$ exists. Second, even if we knew about the convergence toward $\mathcal{S}'$, we would have no general procedure to find out which parts of $\mathcal{S}_i$ will be actually found in $\mathcal{S}'$. This makes it hard to get an intuition for $\mathcal{S}'$ and may be the philosophical reason for Herbrand's rejection of "falsehood in $\mathcal{S}'$" as a meaningful notion. Mathematically, however, we see no justification in Herbrand's rejection of this notion and will explain this in the following.

Herbrand's critical remark concerning the Löwenheim–Skolem Theorem is justified, however, as Löwenheim needs the axiom of choice at two steps in his proof without mentioning this.

1st Step: To show the equivalence of a formula to its outer Skolemized form, Löwenheim's proof requires the full axiom of choice.

2nd Step: For constructing the structure $\mathcal{S}'$, Löwenheim would need König's Lemma, which is a weak form of the axiom of choice.[97]

Contrary to the general perception,[98] there are no essential gaps in Löwenheim's proof, with the exception of the implicit application of the axiom of choice, which was no exception at his time. Indeed, fifteen years later, Gödel still applies the

[94] Actually, the completeness problem is slightly ill defined in [Hilbert and Ackermann, 1928]. Cf. e.g. [Gödel, 1986ff., Vol. I, pp. 44–48].

[95] Cf. [Gödel, 1930].

[96] Letter of Gödel to Heijenoort, dated Aug. 14, 1964. Cf. [Heijenoort, 1971a, p. 510, Note i], [Gödel, 1986ff., Vol. I, p. 51; Vol. V, pp. 315–317].

[97] König's Lemma is Form 10 in [Howard and Rubin, 1998]. This form is even weaker than the well-known Principle of Dependent Choice, namely Form 43 in [Howard and Rubin, 1998]; cf. also [Rubin and Rubin, 1985].

[98] This perception is partly based on the unjustified criticism of Skolem, Herbrand, and Heijenoort.

We are not aware of any negative critique against [Löwenheim, 1915] at the time of publication. [Wang, 1970, p. 27ff.], proof-read by Bernays and Gödel, after being most critical with the proof in [Skolem, 1923b], sees no gaps in Löwenheim's proof, besides the applications of the axiom of choice. The same holds for [Brady, 2000, § 8], sharing expertise in the Peirce–Schröder tradition[40] with Löwenheim.

Let us have a look at the criticism of Skolem, Herbrand, and Heijenoort in detail:

The following statement of Skolem on [Löwenheim, 1915] is confirmed in [Heijenoort, 1971a, p. 230]: *(continued on following page)*

axiom of choice tacitly in the proof of his Completeness Theorem.[99] Moreover, as none of these theorems state any consistency properties, from the point of view of Hilbert's finitism there was no reason to avoid the application of the axiom of choice. Indeed, in the proof of his Completeness Theorem, Gödel "is not interested in avoiding an appeal to the axiom of choice."[100] Thus, again, as we already noted in Item 2 of § 3.1, regarding finitism, Herbrand is more royalist than King Hilbert.

(Note 98, continued from previous page)

> "Er muss also sozusagen einen Umweg über das Nichtabzählbare machen."
> [Skolem, 1923b, p. 220]

> "Thus he must make a detour, so to speak, through the non-denumerable."
> [Heijenoort, 1971a, p. 293, translation by by Stefan Bauer-Mengelberg]

That detour, however, is not an essential part of the proof, but serves for the purpose of illustration only. This is clear from the original paper and also the conclusion of [Badesa, 2004, §§ 3.2 and 3.3].

When Herbrand criticizes Löwenheim's proof, he actually does not criticize the proof as such, but only Löwenheim's semantical notions; even though Herbrand's verbalization suggests the opposite, especially in [Herbrand, 1931, Chapter 2], where Herbrand repeats Löwenheim's reducibility results in finitistic style:

> "Löwenheim [1915] a publié du résultat énoncé dans ce paragraphe une démonstration dont nous avons montré les graves lacunes dans notre travail déjà cité (Chapter 5, § 6.2)." [Herbrand, 1968, p. 187, Note 29]

> "Löwenheim [1915] published a proof of the result stated in this section. In [Herbrand, 1930, Chapter 5, § 6.2], we pointed out that there are serious gaps in his proof." [Herbrand, 1971, p. 237, Note 33, translation by Dreben and Heijenoort]

Heijenoort realized that there is a missing step in Löwenheim's proof:

> "What has to be proved is that, from the assignments thus obtained for all i, there can be formed one assignment such that ΠF is true, that is, $\Pi F = 0$ is false. This Löwenheim does not do." [Heijenoort, 1971a, p. 231]

Except for the principle of choice, however, the missing step is trivial because in Löwenheim's presentation the already fixed part of the assignment is irrelevant for the extension. Indeed, in the "Note to the Second Printing", in the preface of the 2nd printing, Heijenoort partially corrected himself:

> "I am now inclined to think that Löwenheim came closer to König's Lemma than his paper, on the surface, suggests. But a rewriting of my introductory note on that point (p. 231) will have to wait for another occasion." [Heijenoort, 1971a, p. ix]

This correction is easily overlooked because no note was inserted into the actual text.

[99] Cf. [Gödel, 1930].
[100] Cf. [Wang, 1970, p. 24].

The proof of the Löwenheim–Skolem Theorem in [Skolem, 1920] already avoids the axiom of choice in the 1st Step by using *Skolem normal form* instead of Skolemized form.[101]

Moreover, in [Skolem, 1923b], the choices in the 2nd Step of the proof become deterministic, so that no form of the axiom of choice, such as König's Lemma is needed anymore. This is achieved by taking the universe of the structure $\mathcal{S}'$ to be the natural numbers and using the well-ordering of the natural numbers.

THEOREM 9 Löwenheim–Skolem Theorem à la [Skolem, 1923b].
Let Γ be a (finite of infinite) denumerable set of first-order formulas. Assume $\not\models \Gamma$. Without assuming any form of the axiom of choice we can construct a sequence of partial structures $\mathcal{S}_i$ that converges to a structure $\mathcal{S}'$ with a universe which is a subset of the natural numbers such that $\not\models_{\mathcal{S}'} \Gamma$. □

Note that Herbrand does not need any form of the axiom of choice for the following reason: In the 1st Step, Herbrand does not use the semantics of Skolemized forms at all, because Herbrand's Skolem terms are just names for free variables, cf. § 3.10. In the 2nd Step, Herbrand's peculiar notion of "falsehood in an infinite domain" makes any choice superfluous. This is a device which — contrary to what Herbrand wrote — is not really necessary to avoid the axiom of choice, as the above Theorem 9 shows.

In this way, Herbrand came close to proving the completeness of Russell's and Hilbert's calculi for first-order logic, but he did not trust the left arc of the bridge depicted in § 3.10. And thus Gödel proved it first when he submitted his thesis in 1929, in the same year as Herbrand, and the theorem is now called *Gödel's Completeness Theorem* in all textbooks on logic.[102]

It is also interesting to note that Herbrand does not know how to construct a counter-model without using the axiom of choice, as explicitly described in [Skolem, 1923b]. This is — on the one hand — a strong indication that Herbrand was not aware of [Skolem, 1923b].[103] On the other hand, Herbrand names Skolem's Paradox several times and [Skolem, 1923b; 1929] seem to be the only written sources of this at Herbrand's time.[104]

[101] To achieve Skolem normal form, Skolem defines predicates for the subformulas starting with a γ-quantifier, and then rewrites the formula into a prenex form with a first-order $\gamma^*\delta^*$-prefix. Indeed, for the proofs of the versions of the Löwenheim–Skolem Theorem, the Skolemized form (which is used in [Löwenheim, 1915], [Skolem, 1928], and [Herbrand, 1930]) is used neither in [Skolem, 1920] nor in [Skolem, 1923b], which use Skolem normal form instead.

By definition, *Skolemized forms* have a $\delta^*\gamma^*$-prefix with an implicit higher-order δ^*, and *raising* is the dual of Skolemization which produces a $\gamma^*\delta^*$-prefix with a higher-order γ^*, cf. [Miller, 1992]. The *Skolem normal form*, however, has a $\gamma^*\delta^*$-prefix with *first-order* γ^*.

[102] Cf. [Gödel, 1930].

[103] Cf. p.12 of Goldfarb's introduction in [Herbrand, 1971].

[104] Skolem's Paradox is also briefly mentioned in [Neumann, 1925], not as a paradox, however, but as unfavorable conclusions on set theory drawn by Skolem, who wrote about a "peculiar and apparently paradoxical state of affairs", cf. [Heijenoort, 1971a, p. 295].

As Herbrand's Property C and its use of the outer Skolemized form are most similar to the treatment in [Skolem, 1928], it seems likely that Herbrand had read [Skolem, 1928].[105]

3.13 Herbrand's First Proof of the Consistency of Arithmetic

Consider a signature of arithmetic that consists only of zero '$\mathsf{0}$', the successor function 's', and the equality predicate '$=$'. Besides the axioms of equality (equivalence and substitutability), Herbrand considers several subsets of the following axioms:[106]

$$(\mathsf{S}) \qquad P(\mathsf{0}) \ \wedge \ \forall y.\ \big(\ P(y) \Rightarrow P(\mathsf{s}(y))\ \big) \quad \Rightarrow \quad \forall x.\ P(x)$$

$$(\mathsf{nat}_1) \qquad x = \mathsf{0} \ \vee \ \exists y.\ x = \mathsf{s}(y)$$

$$(\mathsf{nat}_2) \qquad \mathsf{s}(x) \neq \mathsf{0}$$

$$(\mathsf{nat}_3) \qquad \mathsf{s}(x) = \mathsf{s}(y) \ \Rightarrow \ x = y$$

$$(\mathsf{nat}_{4+i}) \qquad \mathsf{s}^{i+1}(x) \neq x$$

Axiom (nat_1) together with the well-foundedness of the successor relation 's' specifies the natural numbers up to isomorphism.[107] So do the Dedekind–Peano axioms (nat_2) and (nat_3) together with the Dedekind–Peano axiom of Structural Induction (S), provided that the meta variable P is seen as a universally quantified second-order variable with the standard interpretation.[108]

Of course, Herbrand, the finitist, does not even mention these second-order properties. His discussion is restricted to decidable first-order axiom sets, some of which are infinite due to the inclusion of the infinite sequence (nat_4), (nat_5), (nat_6), ...

As Herbrand's axiom sets are first order, they cannot specify the natural numbers up to isomorphism.[109] But as the model of arithmetic is infinite, Herbrand, the finitist, cannot accept it as part of his proof theory. Actually, he never even mentions the model of arithmetic.

Herbrand shows that (for the poor signature of $\mathsf{0}$, s, and $=$) the first-order theory of the axioms $(\mathsf{nat}_i)_{i\geq 1}$ (i.e. (nat_i) for any positive natural number i) is consistent, complete, and decidable. His constructive proof is elegant, provides a lucid operative understanding of basic arithmetic, and has been included inter

[105] Without giving any justification, Heijenoort assumes, however, that

> "He was not acquainted either, certainly, with [Skolem, 1928]."
>
> [Heijenoort, 1986c, p.112]

[106] The labels are ours, not Herbrand's. Herbrand writes '$x+1$' instead of '$\mathsf{s}(x)$'. To save the axiom of substitutability, Herbrand actually uses the biconditional in (nat_3).

[107] Cf. [Wirth, 2004, § 1.1.3]. This idea goes back to [Pieri, 1907/8].

[108] Cf. e.g. [Andrews, 2002].

[109] For instance, due to the Upward Löwenheim–Skolem–Tarski Theorem. Cf. e.g. [Enderton, 1972].

alia into § 3.1 of [Enderton, 1972], one of the most widely used textbooks on logic. Herbrand's proof has two constructive steps:

1st Step: He shows how to rewrite any formula into an equivalent quantifier-free formula without additional free variables. He proceeds by a special form of quantifier elimination, a technique in the Peirce–Schröder tradition[40] with its first explicit occurrence in [Skolem, 1919].[110]

2nd Step: He shows that the quantifier-free fragment is consistent and decidable and does not depend on the axiom (nat_1). This is achieved with a procedure which rewrites a quantifier-free formula into an equivalent disjunctive normal form without additional free variables. For any quantifier-free formula B, this normal-form procedure satisfies:

$$(\mathsf{nat}_i)_{i\geq 2} \vdash B \quad \text{iff} \quad \text{the normal form of } \neg B \text{ is } 0 \neq 0.$$

This elegant work of Herbrand is hardly discussed in the secondary literature, probably because — as a decidability result — it became obsolete before it was published, due to the analogous result for this theory extended with addition, the so-called *Presburger Arithmetic*, as it is known today. Mojżesz Presburger (1904–1943?)[111] gave his talk on the decidability of his theory with similar techniques on Sept. 24, 1929,[112] five months *after* Herbrand finished his Ph.D. thesis. As Tarski's work on decision methods developed in his 1927/8 lectures in Warszawa also did not appear in print until after World War II,[113] we have to consider this contribution of Herbrand as completely original. Indeed:

> "Dieses gelingt nach einer Methode, welche unabhängig voneinander J. Herbrand and M. Presburger ausgebildet haben. Diese Methode besteht darin, dass man Formeln, welche gebundene Variablen enthalten, ‚Reduzierte' zuordnet, in denen keine gebundenen Variablen mehr auftreten und welche im Sinne der inhaltlichen Deutung jenen Formeln gleichwertig sind."[114]

[110] More precisely, cf. [Skolem, 1919, § 4]. For more information on the subject of quantifier elimination in this context, cf. [Anellis, 1992, p.120f., Note 33], [Wang, 1970, p. 33].

[111] Presburger's true name is Prezburger. He was a student of Alfred Tarski, Jan Łucasiewicz (1878–1956), Kazimierz Ajdukiewicz (1890–1963), and Kazimierz Kuratowski (1896–1980) in Warszawa. He was awarded a master (not a Ph.D.) in mathematics on Oct. 7, 1930. As he was of Jewish origin, it is likely that he died in the Holocaust (Shoah), maybe in 1943. Cf. [Zygmunt, 1991].

[112] Moreover, note that [Presburger, 1930] did not appear in print before 1930. Some citations date [Presburger, 1930] at 1927, 1928, and 1929. There is evidence, however, that these earlier datings are wrong, cf. [Stanisfer, 1984], [Zygmunt, 1991].

[113] Cf. [Stanisfer, 1984], [Tarski, 1951].

[114] Cf. [Hilbert and Bernays, 1968/70, Vol. I, p. 233, note-mark omitted, orthography modernized].
"This is achieved by a method developed independently by J. Herbrand and M. Presburger. The method consists in associating 'reduced forms' to formulas with bound variables, in which no bound variables occur anymore and which are semantically equivalent to the original formulas."

In addition, Herbrand gives a constructive proof that the first-order theories given by the following two axiom sets are identical:

- $(\mathsf{nat}_i)_{i\geq 1}$
- (nat_2), (nat_3), and the first-order instances of (S), provided that (S) is taken as a first-order axiom scheme instead of a second-order axiom.

3.14 Herbrand's Second Proof of the Consistency of Arithmetic

Herbrand's contributions to logic discussed sofar are all published in Herbrand's thesis. In this section, we consider his journal publication [Herbrand, 1932a] as well as some material from Chapter 4 of his thesis.

First, the signature is now enriched to include the recursive functions, cf. § 3.15. Second, the axiom scheme (S) is restricted to just those instances which result from replacing the meta variable P with *quantifier-free* first-order formulas. For this setting, Herbrand again gives a constructive proof of consistency. This proof consists of the following two steps:

1st Step: Herbrand defines recursive functions f_P such that $\mathsf{f}_P(x)$ is the least natural number $y \leq x$ such that $\neg P(y)$ holds, provided that such a y exists, and $\mathsf{0}$ otherwise. The functions f_P are primitive recursive unless the terms substituted for P contain a non-primitive recursive function. These functions imply the instances of (S), rendering them redundant.

This is similar to the effect of Hilbert's 2nd ε-formula:[115]

$$\varepsilon x.\neg P(x) = \mathsf{s}(y) \quad \Rightarrow \quad P(y),$$

Herbrand's procedure, however, is much simpler but only applicable to quantifier-free P.

2nd Step: Consider the universal closures of the axioms of equality, the axioms (nat_2) and (nat_3), and an arbitrary finite subset of the axioms for recursive functions. Take the negation of the conjunction of all these formulas. As all quantified variables of the resulting formula are γ-variables, this is already in Skolemized form. Moreover, for any positive natural number n, it is easy to show that this formula does not have Property C of order n: Indeed, we just have to construct a proper finite substructure of arithmetic which satisfies all the considered axioms for the elements of $\mathcal{T}_n$. Thus, by Herbrand's Fundamental Theorem, consistency is immediate.

[115] Cf. [Hilbert and Bernays, 1968/70, Vol. II, § 2.3, p. 82ff.; Vol. II, Supplement V B, p. 535ff.].

The 2nd step is a prototypical example to demonstrate how Herbrand's Fundamental Theorem helps to answer seemingly non-finitistic semantical questions on infinite structures with the help of infinitely many finite sub-structures. Notice that such a semantical argumentation is finitistically acceptable if and only if the structures are all finite and effectively constructible. And the latter is always the case for Herbrand's work on logic.

As the theory of all recursive functions is sufficiently expressive, there is the question why Herbrand's 2nd consistency proof does not imply the inconsistency of arithmetic by Gödel's second incompleteness theorem? Herbrand explains that we cannot have the theory of *all* total recursive functions because they are not recursively enumerable. More precisely, an evaluation function for an enumerable set of recursive functions cannot be contained in this set by the standard diagonalization argument.

Hilbert's school had failed to prove the consistency of arithmetic, except for the special case that for the axiom (S), the variable x does not occur within the scope of any binder in $P(x)$.[116] But this fragment of arithmetic is actually equivalent to the one considered by Herbrand here.[117] In this sense, Herbrand's result on the consistency of arithmetic was just as strong as the one of the Hilbert school by ε-substitution. Herbrand's means, however, are much simpler.[118]

[116] Cf. [Hilbert and Bernays, 1968/70, Vol. II, § 2.4].

More precisely, Hilbert's school had failed to prove the termination of their first algorithm for computing a valuation of the ε-terms in the 1st and 2nd ε-formulas. This de facto failure was less spectacular but internally more discouraging for Hilbert's programme than Gödel's second incompleteness theorem with its restricted area of application, cf. [Gödel, 1931].

Only after developing a deeper understanding of the notion of a *basic typus* of an ε-term (Grundtypus; introduced in [Neumann, 1927]; called *ε-matrix* in [Scanlon Jr., 1973]) and especially of the independence of its valuation of the valuations of its sub-ordinate ε-expressions, has the problem been resolved: The termination problem was cured in [Ackermann, 1940] with the help of a second algorithm of ε-valuation, terminating within the ordinal number ϵ_0, just as Gentzen's consistency proof for full arithmetic in [Gentzen, 1936; 1938; 1943].

[117] Cf. [Kleene, 1952, p. 474].

[118] Herbrand's results on the consistency of arithmetic have little importance, however, for today's inductive theorem proving because the restrictions on (S) can usually not be met in practice. Herbrand's restrictions on (S) require us to avoid the occurrence of x in the scope of quantifiers in $P(x)$. In practice of inductive theorem proving, this is hardly a problem for the γ-quantifiers, whose bound variables tend to be easily replaceable with witnessing terms. There is a problem, however, with the δ-quantifiers. If we remove the δ-quantifiers, letting their bound δ-variables become free δ-variables, the induction hypothesis typically becomes too weak for proving the induction step. This is because the now free δ-variables do not admit different instantiation in the induction hypotheses and the induction conclusion.

3.15 Foreshadowing Recursive Functions

Herbrand's notion of a recursive function is quite abstract:

A recursive function is given by any new n_i-ary function symbol f_i plus a set of quantifier-free formulas for its specification (which Herbrand calls *the hypotheses*), such that, for any natural numbers $k_1, \ldots, k_{n_i}$, there is a constructive proof of the unique existence of a natural number l such that $\vdash \mathsf{f}_i(\mathsf{s}^{k_1}(0), \ldots, \mathsf{s}^{k_{n_i}}(0)) = \mathsf{s}^l(0)$.

> "On pourra aussi introduire un nombre quelconque de fonctions $\mathsf{f}_i x_1 x_2 \cdots x_{n_i}$ avec des hypothèses telles que:
>
> a) *Elles ne contiennent pas de variables apparentes.*
>
> b) *Considérées intuitionnistiquement,*[5] *elles permettent de faire effectivement le calcul de* $\mathsf{f}_i x_1 x_2 \cdots x_{n_i}$*, pour tout système particulier de nombres; et l'on puisse démontrer intuitionnistiquement que l'on obtient un résultat bien déterminé.* **(Groupe C.)**
>
> [5]Cette expression signifie: traduites en language ordinaire, considérées comme une propriété des entiers, et non comme un pur symbole."
>
> [Herbrand, 1932a, p. 5][119]

In the letter to Gödel dated April 7, 1931, mentioned already in § 2, Herbrand added the requirement that the hypotheses defining f_i contain only function symbols f_j with $j \leq i$, for natural numbers i and j.[120]

Gödel's version of Herbrand's notion of a recursive function is a little different: He speaks of quantifier-free equations instead of quantifier-free formulas and explicitly lists the already known functions, but omits the computability of the functions.

> "If ϕ denotes an unknown function and $\psi_1, \ldots, \psi_k$ are known functions, and if the ψ's and the ϕ are substituted in one another in the most general fashions and certain pairs of the resulting expressions are equated, then, if the resulting set of functional equations has one and only one solution for ϕ, ϕ is a recursive function."[121]

---

[119]Without the colon after "que" and the comma after "$\mathsf{f}_i x_1 x_2 \cdots x_{n_i}$", with a semi-colon instead of a fullstop after "apparentes", and with "langage" instead of "language" also in: [Herbrand, 1968, p. 226f.]

> "We can also introduce any number of functions $\mathsf{f}_i(x_1, x_2, \ldots, x_{n_i})$ together with some hypotheses such that
>
> (a) *The hypotheses contain no bound variables.*
>
> (b) *Considered intuitionistically,*[5] *they make the effective computation of the* $\mathsf{f}_i(x_1, x_2, \ldots, x_{n_i})$ *possible for every given set of numbers, and it is possible to prove intuitionistically that we obtain a well-defined result.* **(Group C.)**
>
> [5]This expression means: translated into ordinary language, considered as a property of integers and not as a mere symbol."

[120]Cf. [Gödel, 1986ff., Vol. V, pp. 14–21], [Sieg, 2005].

[121]Cf. [Gödel, 1934, p. 26], also in: [Gödel, 1986ff., Vol. I, p. 368].

Gödel took this paragraph on page 26 of his 1934 Princeton lectures from the above-mentioned letter from Herbrand to Gödel, which Gödel considered to be lost, but which was rediscovered in February 1986.[122] In 1963, Gödel wrote to Heijenoort:

> "I have never met Herbrand. His suggestion was made in a letter in 1931, and it was formulated *exactly* as on p. 26 of my lecture notes, that is, without any reference to computability. However, since Herbrand was an intuitionist, this definition for him evidently meant that there is a *constructive* proof for the existence and uniqueness of ϕ."[123]

As we have seen, however, Gödel's memory was wrong insofar as he had added the restriction to equations and omitted the computability requirement.

Obviously, Herbrand had a clear idea of our current notion of a total recursive function. Herbrand's characterization, however, just transfers the recursiveness of the meta level to the object level. Such a transfer is of little epistemological value. While there seems to be no way to do much more than such a transfer for consistency of arithmetic in Gödelizable systems (due to Gödel's second incompleteness theorem), it is well possible to do more than that for the notion of recursive functions. Indeed, in the later developments of the theory of term rewriting systems and the today standard recursion theory for total and partial recursive functions, we find constructive definitions and consistency proofs practically useful in programming and inductive theorem proving.[124]

Thus, as suggested by Gödel,[125] we may say that Herbrand *foreshadowed* the notion of a recursive function, although he did not *introduce* it.[126]

---

[122] Cf. [Dawson Jr., 1993].

[123] Letter of Gödel dated April 23, 1963. Cf. [Herbrand, 1971, p. 283], also in: [Heijenoort, 1986c, p.115f.], also in: [Gödel, 1986ff., Vol. V, p. 308].

It is not clear whether Gödel refers to Brouwer's intuitionism or Hilbert's finitism when he calls Herbrand an "intuitionist" here; cf. § 3.1.

And there is more confusion regarding the meaning of two occurrences of the word "intuitionistically" on the page of the above quotation from [Herbrand, 1932a]. Both occurrences carry the same note-mark, probably because Herbrand realized that this time he uses the word with even a another meaning, different from his own standard and different from its meaning for the other occurrence in the same quotation: It neither refers to Brouwer's intuitionism nor to Hilbert's finitism, but actually to the working mathematician's meta level as opposed to the object level of his studies; cf. e.g. [Neumann, 1927, p. 2f.].

[124] Partial recursive functions were introduced in [Kleene, 1938]. For consistency proofs and admissibility conditions for the practical specification of partial recursive functions with positive/negative-conditional term rewriting systems cf. [Wirth, 2008b].

[125] Letter of Gödel to Heijenoort, dated Aug. 14, 1964. Cf. [Heijenoort, 1986c, p.115f.].

[126] Cf. [Heijenoort, 1986c, p.115ff.] for more on this. Moreover, note that a general definition of (total) recursive functions was not required for Herbrand's second consistency proof because Herbrand's function f_P of § 3.14 is actually a *primitive* recursive one, unless P contains a non-primitive recursive function.

4 INFLUENCE

Despite his very short life, Herbrand's contributions were of great significance at his time and they had a strong impact on the work by others later in mathematics, proof theory, computer science, and artificial intelligence. Even today the name "Herbrand" can be found astonishingly often in research papers in fields that did not even exist at his time.[127]

We sketch a list of topics that were influenced by Herbrand's work.

4.1 Proof Theory

Dirk van Dalen (born 1932) begins his review on [Herbrand, 1971] as follows:

> "Much of the logical activity in the first half of this century was inspired by Hilbert's programme, which contained, besides fundamental reflections on the nature of mathematics, a number of clear-cut problems for technically gifted people. In particular the quest for so-called "consistency proofs" was taken up by quite a number of logicians. Among those, two men can be singled out for their imaginative approach to logic and mathematics: Jacques Herbrand and Gerhard Gentzen. Their contributions to this specific area of logic, called "proof theory" (Beweistheorie) following Hilbert, are so fundamental that one easily recognizes their stamp in modern proof theory."[128]

Dalen continues:

> "When we realize that Herbrand's activity in logic took place in just a few years, we cannot but recognize him as a giant of proof theory. He discovered an extremely powerful theorem and experimented with it in proof theory. It is fruitless to speculate on the possible course Herbrand would have chosen, had he not died prematurely; a consistency proof for full arithmetic would have been within his reach."[129]

[127] To wit, the search in any online library (e.g. citeseer) reveals that astonishingly many authors dedicate parts of their work directly to Jacques Herbrand. A "Google Scholar" search gives a little less than ten thousand hits and the phrases we find by such an experiment include: Herbrand agent language, Herbrand analyses, Herbrand automata, Herbrand base, Herbrand complexity, Herbrand constraints, Herbrand disjunctions, Herbrand entailment, Herbrand equalities, Herbrand expansion, Herbrand's Fundamental Theorem, Herbrand functions, Herbrand–Gentzen theorem, Herbrand interpretation, Herbrand–Kleene universe, Herbrand model, Herbrand normal forms, Herbrand procedures, Herbrand quotient, Herbrand realizations, Herbrand semantics, Herbrand strategies, Herbrand terms, Herbrand–Ribet theorem, Herbrand's theorem, Herbrand theory, Herbrand universe. Whether and to what extend these references to Herbrand are justified is sometimes open for debate. This list shows, however, that in addition to the foundational importance of his work at the time, his insights still have an impact on research even at the present time. Herbrand's name is therefore not only frequently mentioned among the most important mathematicians and logicians of the 20th century but also among the pioneers of modern computer science and artificial intelligence.

[128] Cf. [Dalen, 1974, p. 544].

[129] Cf. [Dalen, 1974, p. 548].

The major thesis of [Anellis, 1991] is that, building on the Löwenheim–Skolem Theorem, it was Herbrand's work in elaborating Hilbert's concept of "being a proof" that gave rise to the development of the variety of first-order calculi in the 1930s, such as the ones of the Hilbert school, and such as Natural Deduction and Sequent calculi in [Gentzen, 1935].

As already indicated in § 3.9, Herbrand's Fundamental Theorem has directly influenced Bernays' work on proof theory.

The main inspiration in the *unwinding programme*,[130] which — to save the merits of proof theory — Georg Kreisel (born 1923) suggested as a replacement for Hilbert's failed programme, is Herbrand's Fundamental Theorem, especially for Kreisel's notion of a *recursive interpretation* of a logic calculus in another one,[131] such as given by Herbrand's Fundamental Theorem for his first-order calculus in the sentential tautologies over the language enriched with Skolem functions.

Herbrand's approach to consistency proofs, as we have sketched in §§ 3.13 and 3.14, has a semantical flavor and is inspired by Hilbert's evaluation method of ε-substitution, whereas it avoids the dependence on Hilbert's ε-calculus. The main idea (cf. § 3.14) is to replace the induction axiom by recursive functions of finitistic character. Herbrand's approach is in contrast to the purely syntactical style of Gentzen[132] and Schütte[133] in which semantical interpretation plays no rôle.

So-called *Herbrand-style consistency proofs* follow Herbrand's idea of constructing finite sub-models to imply consistency by Herbrand's Fundamental Theorem. During the early 1970s, this technique was used by Thomas M. Scanlon Jr. (born 1940) in collaboration with Dreben and Goldfarb.[134] These consistency proofs for arithmetic roughly follow Ackermann's previous proof,[135] but they apply Herbrand's Fundamental Theorem in advance and consider Skolemized form instead of Hilbert's ε-terms.[136]

---

[130] Cf. [Kreisel, 1951; 1952; 1958; 1982], and do not miss the discussion in [Feferman, 1996]!

> "*To determine the constructive (recursive) content or the constructive equivalent of the non-constructive concepts and theorems used in mathematics*, particularly arithmetic and analysis." [Kreisel, 1958, p. 155]

[131] Cf. [Kreisel, 1958, p. 160] and [Feferman, 1996, p. 259f.].

[132] Cf. [Gentzen, 1936; 1938; 1943].

[133] Cf. [Schütte, 1960].

[134] Cf. [Dreben and Denton, 1970], [Scanlon Jr., 1973], [Goldfarb and Scanlon Jr., 1974].

[135] Cf. [Ackermann, 1940].

[136] Contrary to the proofs of [Dreben and Denton, 1970] and [Ackermann, 1940], the proof of [Scanlon Jr., 1973], which is otherwise similar to the proof of [Dreben and Denton, 1970], admits the inclusion of induction axioms over *any* recursive well-ordering on the natural numbers:

By an application of Herbrand's Fundamental Theorem, from a given derivation of an inconsistency, we can compute a positive natural number n such that Property C of order n holds. Therefore, in his analog of Hilbert's and Ackermann's ε-substitution method, Scanlon can effectively pick a minimal counterexample on the champ fini $\mathcal{T}_n$ from a given critical counterexample, even if this neither has a direct predecessor nor a finite initial segment.

This result was then further generalized in [Goldfarb and Scanlon Jr., 1974] to ω-consistency of arithmetic.

Gentzen's and Herbrand's insight on Cut and *modus ponens* elimination and the existence of normal form derivations with mid-sequents had a strong influence on William Craig's work on interpolation.[137] The impact of *Craig's Interpolation Theorem* to various disciplines in turn has recently been discussed at the Interpolations Conference in Honor of William Craig in May 2007.[138]

4.2 Recursive Functions and Gödel's Second Incompleteness Theorem

As already discussed in detail in § 3.15, in his 1934 Princeton lectures, Gödel introduced the notion of *(general) recursive functions* and mentioned that this notion had been proposed to him in the letter from Herbrand mentioned already in § 2. This letter, however, seems to have had more influence on Gödel's thinking, namely on the consequences of Gödel's second incompleteness theorem:

> "Nowhere in the correspondence does the issue of *general* computability arise. Herbrand's discussion, in particular, is solely trying to explore the limits of consistency proofs that are imposed by the second theorem. Gödel's response also focuses on that very topic. It seems that he subsequently developed a more critical perspective on the very character and generality of this theorem." [Sieg, 2005, p. 180]

Sieg [2005] argues that the letter of Herbrand to Gödel caused a change of Gödel's perception of the impact of his own second incompleteness theorem on Hilbert's programme: Initially Gödel did assert that it would not contradict Hilbert's viewpoint. Influenced by Herbrand's letter, however, he accepted the more critical opinion of Herbrand on this matter.

4.3 Algebra and Ring Theory

In 1930–1931, within a few months, Herbrand wrote several papers on algebra and ring theory. During his visit to Germany he met and briefly worked on this topic with Noether, Hasse, and Artin; cf. § 2. He contributed several new theorems of his own and simplified proofs of results by Leopold Kronecker (1823–1891), Heinrich Weber (1842–1913), Teiji Takagi (1875–1960), Hilbert, and Artin, thereby generalizing some of these results. The Herbrand–Ribet Theorem is a result on the class number of certain number fields and it strengthens Kummer's convergence criterion; cf. Figure 1.

There is no direct connection between Herbrand's work on logic and his work on algebra. Useful applications of proof theory to mathematics are very rare. Kreisel's "unwinding" of Artin's proof of Artin's Theorem into a constructive proof seems to be one of the few exceptions.[139]

[137] Cf. [Craig, 1957a; Craig, 1957b].
[138] Cf. `http://sophos.berkeley.edu/interpolations/`.

"An odd prime p is called irregular if the class number of the field $\mathbf{Q}(\mu_p)$ is divisible by p (μ_p being, as usual, the group of p-th roots of unity). According to Kummer's criterion, p is irregular if and only if there exists an even integer k with $2 \leq k \leq p-3$ such that p divides (the numerator of) k-th Bernoulli number B_k, given by the expansion

$$\frac{t}{e^t - 1} + \frac{t}{2} - 1 = \sum_{n \geq 2} \frac{B_n}{n!} t^n.$$

The purpose of this paper is to strengthen Kummer's criterion.
Let A be the ideal class group of $\mathbf{Q}(\mu_p)$, and let C be the F_p-vector space A/A^p. The Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on C through its quotient $\Delta = \mathrm{Gal}(\mathbf{Q}(\mu_p)/\mathbf{Q})$. Since all characters of Δ with values in $\overline{\mathbf{F}}_p^*$ are powers of the standard character

$$\chi : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \Delta \xrightarrow{\sim} \mathbf{F}_p^*$$

giving the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on μ_p, the vector space C has a canonical decomposition

$$C = \bigoplus_{i \bmod (p-1)} C(\chi^i),$$

where

$$C(\chi^i) = \{c \in C | \sigma c = \chi^i(\sigma)\, c \text{ for all } \sigma \in \Delta\}.$$

(1.1) **Main Theorem.**
Let k be even, $2 \leq k \leq p-3$.
Then $p|B_k$ if and only if $C(\chi^{1-k}) \neq 0$
In fact, the statement that $C(\chi^{1-k}) \neq 0$ implies $p|B_k$ is well known [Herbrand, 1932b, Th. 3]"
[Ribet, 1976, Theorem 1.1, p. 151]

Figure 1. Herbrand–Ribet Theorem

[139] With Artin's Theorem we mean:

> "Eine rationale Funktion $F(x_1, x_2, \cdots, x_n)$ von n Veränderlichen heiße definit, wenn sie für kein reelles Wertsystem der x_i negative Werte annimmt."
> [Artin, 1927, p. 100]

> "Satz 4: Es sei R ein reeller Zahlkörper, der sich nur auf eine Weise ordnen lässt, wie zum Beispiel der Körper der rationalen Zahlen, oder der aller reellen algebraischen Zahlen oder der aller reellen Zahlen.
> Dann ist jede rationale definite Funktion von $x_1, x_2, \cdots, x_n$ mit Koëffizienten aus R Summe von Quadraten von rationalen Funktionen der x_i mit Koëffizienten aus R."
> [Artin, 1927, p. 109, modernized orthography]

Cf. [Delzell, 1996] for the unwinding of Artin's proof of Artin's Theorem.
Cf. [Feferman, 1996] for a discussion of the application of proof theory to mathematics in general.

4.4 Automated Deduction

In the last fifty years the field of automated deduction, or automated reasoning as it is more generally called today, has come a long way: modern deduction systems are among the most sophisticated and complex human artefacts we have, they can routinely search spaces of several million formulas to find a proof. Automated theorem proving systems have solved open mathematical problems and these deduction engines are used nowadays in many subareas of computer science and artificial intelligence, including software development and verification as well as security analysis. The application in industrial software and hardware development is now standard practice in most high quality products. The two volume handbook [Robinson and Voronkow, 2001] gives a good impression of the state of the art today.

Herbrand's work inspired the development of the first computer programs for automated deduction and mechanical theorem proving, for the following reason: The actual test for Herbrand's Property C is very mechanical in nature and thus can be carried out on a computer, resulting in a mechanical semi-decision procedure for any mathematical theorem! This insight, first articulated in the 1950s, turned out to be most influential in automated reasoning, artificial intelligence, and computer science.

Let us recapitulate the general idea as it is common now in most monographs and introductory textbooks on automated theorem proving.

Suppose we are given a conjecture A. Let F be its (validity) Skolemized form. We then eliminate all quantifiers in F, and the result is a quantifier-free formula E. We now have to show that the Herbrand disjunction over some possible values of the free variables of E is valid. See also Example 4 in § 3.8.

How do we find these values? Well, we do not actually have these "objects in the domain", but we can use their *names*, i.e. we take all terms from the *Herbrand universe* and substitute them in a systematic way into the variables and wait what happens: every substituted formula obtained that way is sentential, so we can just check whether their disjunction is valid or not with one of the many available decision procedures for sentential logic. If it is valid, we are done: the original formula A must be valid by Herbrand's Fundamental Theorem. If the disjunction of instantiated sentential formulas turns out to be invalid, well, then bad luck and we continue the process of substituting terms from the *Herbrand universe*.

This process must terminate, if indeed the original theorem is valid. But what happens if the given conjecture is in fact not a theorem? Well, in that case the process will either run forever or sometimes, if we are lucky, we can nevertheless show this to be the case.

In the following we will present these general ideas a little more technically.[140]

[140] Standard textbooks covering the early period of automated deduction are [Chang and Lee, 1973] and [Loveland, 1978]. Chang and Lee [1973] present this and other algorithms in more detail and rigor.

Arithmetic provided a testbed for the first automated theorem proving program: In 1954 the program of Martin Davis (born 1928) proved the exciting theorem that the sum of two even numbers is again even. This date is still considered a hallmark and was used as the date for the 50th anniversary of automated reasoning in 2004. The system that proved this and other theorems was based on Presburger Arithmetic, a decidable fragment of first-order logic, mentioned already in § 3.13.

Another approach, based directly on Herbrand's ideas, was tried by Paul C. Gilmore (born 1925).[141] His program worked as follows: A preprocessor generated the *Herbrand disjunction* in the following sense. The formula E contains finitely many constant and function symbols, which are used to systematically generate the *Herbrand universe* for this set; say

$$a, b, f(a,a), f(a,b), f(b,a), f(b,b), g(a,a), g(a,b), g(b,a), g(b,b), f(a, f(a,a)), \ldots$$

for the constants a, b and the binary function symbols f, g. The terms of this universe were enumerated and systematically substituted for the variables in E such that the program generates a sequence of propositional formulas $E\sigma_1, E\sigma_2, \ldots$ where σ_1, σ_2 are the substitutions. Now each of these sets can be checked for truth-functional validity, for which Gilmore used the "multiplication method". This method computes the conjunctive normal form[142] and checks the individual elements of this conjunction in turn: If any element contains the disjunction of an atom and its negation, it must be true and hence can be removed from the overall conjunction. As soon as all disjunctions have been removed, the theorem is proved — else it goes on forever.

This method is not particularly efficient and served to prove a few very simple theorems only. Such algorithms became known as *British Museum Algorithms*. That name was originally justified as follows:

> "Thus we reflect the basic nature of theorem proving; that is, its nature prior to building up sophisticated proof techniques. We will call this algorithm the British Museum Algorithm, in recognition of the supposed originators of this type." [Newell *et al.*, 1957]

The name has found several more poplular explanations since. The nicest is the following: If monkeys are placed in front of typewriters and they type in a guaranteed random fashion, they will reproduce all the books of the library of the British Museum, provided they could type long enough.

[141] Cf. [Gilmore, 1960]. The idea is actually due to Löwenheim and Skolem besides Herbrand, as discussed in detail in § 3.

[142] Actually, the historic development of automated theorem proving did not follow Herbrand but Skolem in choosing the duality validity–unsatisfiablity. Thus, instead of proving validity of a conjecture, the task was to show unsatisfiablity of the negated conjecture. For this reason, Gilmore actually used the *disjunctive* normal form here.

A few months later Martin Davis and Hilary Putnam (born 1926) experimented with a better idea, where the multiplication method is replaced by what is now known as the *Davis–Putnam procedure.*[143] It works as follows: the initial formula E is transformed (once and for all) into disjunctive normal form and then the variables are systematically replaced by terms from the Herbrand universe as before. But now the truth-functional check is replaced with a very effective procedure, which constituted a huge improvement and is still used today in many applications involving propositional logic.[144] However, the most cumbersome aspect remained: the systematic but blind replacement of variables by terms from the Herbrand universe. Could we not find these replacements in a more goal-directed and intelligent way?

The first step in that direction was done by Davis [1963] in a method called linked conjuncts, where the substitution was cleverly chosen, so that it generated the desired tautologies more directly. And this idea finally led to the seminal resolution principle discovered by J. Alan Robinson (born 1930), which dominated the field ever since.[145]

This technique — called a *machine-oriented* logic by Robinson — dispenses with the systematic replacement from the Herbrand universe altogether and finds the proper substitutions more directly by an ingenious combination of Cut and a unification algorithm. It works as follows: first transform the formula E into disjunctive normal form, i.e. into a disjunctive set of conjunctions. Now suppose that the following elements are in this disjunctive set:

$$K_1 \wedge \ldots \wedge K_m \wedge L \quad \text{and} \quad \neg L \wedge M_1 \wedge \ldots \wedge M_n.$$

Then we can add their *resolvent*

$$K_1 \wedge \ldots \wedge K_m \wedge M_1 \wedge \ldots \wedge M_n$$

to this disjunction, simply because one of the previous two must be true if the resolvent is true.

Now suppose that the literals L and $\neg L$ are not already complementary because they still contain variables, for example such as $P(x, f(a, y))$ and $\neg P(a, f(z, b))$. It is easy to see, that these two atoms can be made equal, if we substitute a for the variables x and z and the constant b for the variable y. The most important aspect of the resolution principle is that this substitution can be computed by an algorithm, which is called *unification*. Moreover, there is always *at most one* (up to renaming) most general substitution which unifies two atoms, and this single unifier stands for the potentially infinitely many instances from the Herbrand universe that would be generated otherwise.

Robinson's original unification algorithm is exponential in time and space. The race for the fastest *unification algorithm* lasted more than a quarter of a century and resulted in a linear algorithm and unification theory became a (small) subfield of computer science, artificial intelligence, logic, and universal algebra.[146]

[143] Cf. [Davis and Putnam, 1960].
[144] Cf. The international SAT Competitions web page `http://www.satcompetition.org/`.
[145] Cf. [Robinson, 1965].
[146] Cf. [Siekmann, 1989] for a survey.

Unification theory had its heyday in the late 1980s, when the Japanese challenged the Western economies with the "Fifth Generation Computer Programme" which was based among others on logical programming languages. The processors of these machines realized an ultrafast unification algorithm cast in silicon, whose performance was measured not in MIPS (machine instructions per second), as with standard computers, but in LIPS (logical inferences per second, which amounts to the number of unifications per second). A myriad of computing machinery was built in special hardware or software on these new concepts and most industrial countries even founded their own research laboratories to counteract the Japanese challenge.[147]

Interestingly, Jacques Herbrand had seen the concept of a unifying substitution and an algorithm already in his thesis in 1929. Here is his account in the original French idiom:

> "1^0. Si une des égalités à satisfaire égale une variable restreinte x à un autre individu; ou bien cet individu contient x, et on ne peut y satisfaire; ou bien il ne contient pas x; cette égalité sera alors une des égalités normales cherchées; et on remplacera x par cette fonction dans les autres égalités à satisfaire.
>
> 2^0. Si une des égalités à satisfaire égale une variable générale à un autre individu, qui ne soit pas une variable restreinte, il est impossible d'y satisfaire.
>
> 3^0. Si une des égalités à satisfaire égale
> $$f_1(\varphi_1, \varphi_2, \ldots, \varphi_n) \text{ à } f_2(\psi_1, \psi_2, \ldots, \psi_m),$$
> ou bien les fonctions élémentaires f_1 et f_2 sont différentes, auquel cas il es impossible d'y satisfaire; ou bien les fonctions f_1 et f_2 sont les mêmes; auquel cas on remplace l'égalité par celles obtenues en égalant φ_i à ψ_i."[148]

[147] Such as the European Computer-Industry Research Center (ECRC), which was supported by the French company Bull, the German Siemens company, and the British company ICL.

[148] Cf. [Herbrand, 1930, p. 96f.]. Note that the reprint in [Herbrand, 1968, p.124] has "associées" instead of "normales", which is not an improvement, and the English translation of [Herbrand, 1971, p.148] is based on this contorted version.

> "1. If one of the equations to be satisfied equates a γ-variable x^γ to another term; either this term contains x^γ, and then the equation cannot be satisfied; or else the term does not contain x^γ, and then the equation will be one of the normal-form equations we are looking for, and we replace x^γ by the term in the other equations to be satisfied.
>
> 2. If one of the equations to be satisfied equates a δ-variable to another term that is not a γ-variable, the equation cannot be satisfied.
>
> 3. If one of the equations to be satisfied equates $f_1(\varphi_1, \varphi_2, \ldots, \varphi_n)$ to $f_2(\psi_1, \psi_2, \ldots, \psi_m)$; either the function symbols f_1 and f_2 are different, and then the equation cannot be satisfied; or they are the same, and then we replace this equation with those that equate φ_i to ψ_i."

ACKNOWLEGDGEMENTS

We would like to thank our second readers Irving H. Anellis, Paolo Bussotti, and Lawrence C. Paulson for their numerous, extensive, deep, and most valuable and helpful suggestions for improvement of previous versions of this paper. As always, the second readers are, of course, not responsible for the remaining weaknesses.

Furthermore, we would like to thank John W. Dawson Jr., Christian G. Fermüller, Bernhard Gramlich, Andreas Hallmann, Georg Ch. Moser, Peter Roquette, Andrea Sattler-Klein, James T. Smith, and Magdalena Wolska for substantial help.

BIBLIOGRAPHY OF JACQUES HERBRAND

To our best knowledge, the following bibliography of Jacques Herbrand (Jr.) is complete, with the exception of his letters in [Gödel, 1986ff., Vol. V, pp. 3–25] and [Roquette, 2008].
Note that for labels of the form [Herbrand ...], we have maintained the standard labeling as introduced by Jean van Heijenoort in [Herbrand, 1968; 1971] and [Heijenoort, 1971a] as far as appropriate. For instance, Herbrand's thesis is cited as [Herbrand, 1930] and not as [Herbrand, 1930a], which would be the standard style of this handbook and which is in general preferable due to the additional redundancy.

[Herbrand and Chevalley, 1931] Jacques Herbrand and Claude Chevalley. Nouvelle démonstration du théorème d'existence en théorie du corps de classes. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences (Paris)*, 193:814–815, 1931.

[Herbrand, 1928] Jacques Herbrand. Sur la théorie de la démonstration. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences (Paris)*, 186:1274–1276, 1928. Also in [Herbrand, 1968, pp. 21–23]. Annotated English translation *On Proof Theory* by Warren Goldfarb with an extended note by Burton Dreben and Goldfarb in [Herbrand, 1971, pp. 29–34].

[Herbrand, 1929] Jacques Herbrand. Non-contradiction des axiomes arithmétiques. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences (Paris)*, 188:303–304, 1929. Also in [Herbrand, 1968, pp. 25–26]. Annotated English translation *The Consistency of the Axioms of Arithmetic* by Warren Goldfarb in [Herbrand, 1971, pp. 35–37]. *(An abstract of [Herbrand, 1930, Chapter 4]).*

[Herbrand, 1929a] Jacques Herbrand. Sur quelques propriétés des propositions vraies et leurs applications. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences (Paris)*, 188:1076–1078, 1929. Also in [Herbrand, 1968, pp. 27–29]. Annotated English translation *On several properties of true propositions and their applications* by Warren Goldfarb in [Herbrand, 1971, pp. 38–40]. *(An abstract of [Herbrand, 1930, Chapter 5, § 6]).*

[Herbrand, 1929b] Jacques Herbrand. Sur le probléme fondamental des mathématiques. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences (Paris)*, 189:554–556, 1929. Also in [Herbrand, 1968, pp. 31–33]. Annotated English translation *On the Fundamental Problem of Mathematics* by Warren Goldfarb in [Herbrand, 1971, pp. 41–43]. *(An abstract of [Herbrand, 1931]).*

[Herbrand, 1930] Jacques Herbrand. *Recherches sur la Théorie de la Démonstration.* PhD thesis, Université de Paris, 1930. Thèses présentées à la faculté des Sciences de Paris pour obtenir le grade de docteur ès sciences mathématiques — 1re thèse: Recherches sur la théorie de la démonstration — 2me thèse: Propositions données par la faculté, Les équations de Fredholm — Soutenues le 1930 devant la commission d'examen — Président: M. Vessiot, Examinateurs: MM. Denjoy, Frechet — Vu et approuvé, Paris, le 20 Juin 1929, Le doyen de la faculté des Sciences, C. Maurain — Vu et permis d'imprimer, Paris, le 20 Juin 1929, Le recteur de l'Academie de Paris, S. Charlety — No. d'ordre 2121, Série A, No. de Série 1252 — Imprimerie J. Dziewulski, Varsovie — Univ. de Paris. Also in Prace Towarzystwa Naukowego Warszawskiego, Wydział III Nauk Matematyczno-Fizychnych, Nr. 33, Warszawa. Also in [Herbrand, 1968, pp. 35–153]. Annotated English translation *Investigations in Proof Theory* by Warren Goldfarb (Chapters 1–4) and Burton Dreben and Jean van Heijenoort (Chapter 5) with a brief introduction by Goldfarb and extended notes by Goldfarb (Notes A–C, K–M, O), Dreben (Notes F–I), Dreben and Goldfarb (Notes D, J, and N), and Dreben, George Huff, and Theodore Hailperin (Note E) in [Herbrand, 1971, pp. 44–202]. English translation of § 5 with a different introduction by Heijenoort and some additional extended notes by Dreben also in [Heijenoort, 1971a, pp. 525–581]. *(Herbrand's Ph.D. thesis, his cardinal work, dated April 14, 1929; submitted at the Univ. of Paris; defended at the Sorbonne June 11, 1930; printed in Warsaw, 1930.)*

[Herbrand, 1930a] Jacques Herbrand. Les bases de la logique hilbertienne. *Revue de Métaphysique et de Morale*, 37:243–255, 1930. Also in [Herbrand, 1968, pp. 155–166]. Annotated English translation *The Principles of Hilbert's Logic* in [Herbrand, 1971, pp. 203–214]. *(Herbrand's personal view on meta-mathematics and Hilbert's programme).*

[Herbrand, 1930b] Jacques Herbrand. Nouvelle démonstration et généralisation d'un théorème de Minkowski. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences (Paris)*, 191:1282–1285, 1930.

[Herbrand, 1931] Jacques Herbrand. Sur le problème fondamental de la logique mathématique. *Revue de Métaphysique et de Morale*, 24:12–56, 1931. Also in [Herbrand, 1968, pp. 167–207]. Annotated English translation *On the Fundamental Problem of Mathematical Logic* in [Herbrand, 1971, pp. 215–271].

[Herbrand, 1931a] Jacques Herbrand. Note non signeé sur [Herbrand, 1930]. *Annales de l'Université de Paris*, 6:186–189, 1931. Also in [Herbrand, 1968, pp. 209–214]. English translation *Unsigned Note on Herbrand's Thesis written by Herbrand himself* in [Herbrand, 1971, pp. 272–276]. *(The philosophy of metamathematics in the Hilbert school (roughly following [Hilbert, 1928]) and the positioning of [Herbrand, 1930] in it are sketched for non-experts).*

[Herbrand, 1931b] Inconsistent label replaced with [Herbrand, 1931d] or [Herbrand, 1932a].

[Herbrand, 1931c] Inconsistent label replaced with [Herbrand, 1932a].

[Herbrand, 1931d] Jacques Herbrand. Notice pour Jacques Hadamard. 1931. In [Herbrand, 1968, pp. 215–219]. With annotations and a brief introduction by Jean van Heijenoort. English translation *Note for Jacques Hadamard* by Warren Goldfarb in [Herbrand, 1971, pp. 277–281]. *(Abstracts for non-experts of Herbrand's mathematical contributions addressed to his presenter to the Academy of Sciences).*

[Herbrand, 1931e] Jacques Herbrand. Sur la théorie des corps de nombres de degré infini. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences (Paris)*, 193:594, 1931. *(Abstract of [Herbrand, 1932d; 1933]).*

[Herbrand, 1931f] Jacques Herbrand. Sur la théorie des groupes de décomposition, d'inertie et de ramication. *J. de Math. Pures et Appliquées (J. Liouville)*, 9:481–498, 1931.

[Herbrand, 1932a] Jacques Herbrand. Sur la non-contradiction de l'Arithmetique. *J. für die reine und angewandte Mathematik (Crellesches J.)*, 166:1–8, 1932. Received July 27, 1931. Without Hasse's obituary also in [Herbrand, 1968, pp. 221–232]. Annotated English translation *On the Consistency of Arithmetic* in [Heijenoort, 1971a, pp. 618–628]. Annotated English translation also in [Herbrand, 1971], pp. 282–298. *(Consistency of arithmetic, foreshadowing of (total) recursive functions, discussion of [Gödel, 1931]).*

[Herbrand, 1932b] Jacques Herbrand. Sur les classes des corps circulaires. *J. de Math. Pures et Appliquées (J. Liouville)*, 9:417–441, 1932.

[Herbrand, 1932c] Jacques Herbrand. Sur les théorèmes du genre principal et des idéaux principaux. *Abhandlungen aus dem mathematischen Seminar der Univ. Hamburg*, 9:84–92, 1932.

[Herbrand, 1932d] Jacques Herbrand. Théorie arithmétique des corps de nombres de degré infini — I. Extensions algébriques finies de corps infinis. *Mathematische Annalen*, 106:473–501, 1932.

[Herbrand, 1932e] Jacques Herbrand. Une propriété du discriminant des corps algébriques. *Annales Scientifiques d l'École Normale Supérieure*, 49:105–112, 1932.

[Herbrand, 1932f] Jacques Herbrand. Zur Theorie der algebraischen Funktionen. *Mathematische Annalen*, 106:502, 1932.

[Herbrand, 1933] Jacques Herbrand. Théorie arithmétique des corps de nombres de degré infini — II. Extensions algébriques de degré infini. *Mathematische Annalen*, 108:699–717, 1933. Ed. and with a last paragraph by Claude Chevalley.

[Herbrand, 1936] Jacques Herbrand. *Le Développement Moderne de la Théorie des Corps Algébriques — Corps de classes et lois de réciprocité.* Mémorial des Sciences Mathématiques, Fascicule LXXV. Gauthier-Villars, Paris, 1936. Ed. and with an appendix by Claude Chevalley.

[Herbrand, 1968] Jacques Herbrand. *Écrits Logiques.* Presses Universitaires de France, Paris, 1968. Ed. by Jean van Heijenoort. English translation is [Herbrand, 1971].

[Herbrand, 1971] Jacques Herbrand. *Logical Writings.* Harvard Univ. Press, 1971. Ed. by Warren Goldfarb. Translation of [Herbrand, 1968] with additional annotations, brief introductions, and extended notes by Goldfarb, Burton Dreben, and Jean van Heijenoort. *(Still the best source on Herbrand's logical writings today).*

REFERENCES

[Ackermann, 1925] Wilhelm Ackermann. Begründung des ‚tertium non datur' mittels der Hilbertschen Theorie der Widerspruchsfreiheit. *Mathematische Zeitschrift*, 93:1–36, 1925. Received March 30, 1924.

[Ackermann, 1928] Wilhelm Ackermann. Über die Erfüllbarkeit gewisser Zählausdrücke. *Mathematische Annalen*, 100:638–649, 1928. Received Feb. 9, 1928.

[Ackermann, 1940] Wilhelm Ackermann. Zur Widerspruchsfreiheit der Zahlentheorie. *Mathematische Annalen*, 117:163–194, 1940. Received Aug. 15, 1939.

[Andrews, 2002] Peter B. Andrews. *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*. Academic Press (Elsevier), 2002. 2nd ed. (1st ed. 1986).

[Andrews, 2003] Peter B. Andrews. Herbrand Award acceptance speech. *J. Automated Reasoning*, 31:169–187, 2003.

[Anellis, 1991] Irving H. Anellis. The Löwenheim–Skolem Theorem, theories of quantification, and proof theory. 1991. In [Drucker, 1991, pp. 71–83].

[Anellis, 1992] Irving H. Anellis. *Logic and Its History in the Work and Writings of Jean van Heijenoort*. Modern Logic Publ., Ames (IA), 1992.

[Anellis, 2006] Irving H. Anellis. Gödel and the naming of Skolem functions. In *Sovremennaja logika: problemy teorii, istorii i primenenija v nauke. Materialy IX Obshcherossijskoj nauchnoj konferencii. Sankt-Peterburg, June 22–24, 2006*, pages 410–413. Izdatelśtvo Sankt Peterburgskogo Universiteta, St. Peterburg, 2006. ISBN 5-288-03998-4.

[Artin and Schreier, 1927] Emil Artin and Otto Schreier. Algebraische Konstruktion reeller Körper. *Abhandlungen aus dem mathematischen Seminar der Univ. Hamburg*, 5:85–99, 1927. Hamburg, Mathematisches Seminar, June 1926. Facsimile also in [Artin, 1965, pp. 258–272].

[Artin, 1927] Emil Artin. Über die Zerlegung definiter Funktionen in Quadrate. *Abhandlungen aus dem mathematischen Seminar der Univ. Hamburg*, 5:100–115, 1927. Hamburg, Mathematisches Seminar, June 1926. Facsimile also in [Artin, 1965, pp. 273–288].

[Artin, 1965] Emil Artin. *Collected Papers*. Addison-Wesley, 1965. Ed. by Serge Lang and John T. Tate. Reprinted 1982 by Springer.

[Autexier, 2005] Serge Autexier. The CORE calculus. In *20th Int. Conf. on Automated Deduction, Tallinn, 2005*, number 3632 in Lecture Notes in Artificial Intelligence, pages 84–98. Springer, 2005.

[Baaz and Fermüller, 1995] Matthias Baaz and Christian G. Fermüller. Non-elementary speedups between different versions of tableaux. In *Int. Conf. on Tableaus and Related Methods, TABLEAUX'95*, number 918 in Lecture Notes in Artificial Intelligence, pages 217–230. Springer, 1995.

[Baaz and Leitsch, 1995] Matthias Baaz and Alexander Leitsch. Methods of functional extension. *Collegium Logicum — Annals of the Kurt Gödel Society*, 1:87–122, 1995.

[Badesa, 2004] Calixto Badesa. *The Birth of Model Theory — Löwenheim's Theorem in the Frame of the Theory of Relatives*. Princeton Univ. Press, 2004. Translated by Michaël Maudsley.

[Behmann, 1922] Heinrich Behmann. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Mathematische Annalen*, 86:163–229, 1922.

[Berka and Kreiser, 1973] Karel Berka and Lothar Kreiser, editors. *Logik-Texte – Kommentierte Auswahl zur Geschichte der modernen Logik*. Akademie-Verlag, Berlin, 1973. 2nd rev. ed. (1st ed. 1971; 4th rev. rev. ed. 1986).

[Bernays and Schönfinkel, 1928] Paul Bernays and Moses Schönfinkel. Zum Entscheidungsproblem der mathematischen Logik. *Mathematische Annalen*, 99:342–372, 1928.

[Bernays, 1928] Paul Bernays. Zusatz zu Hilberts Vortrag „Die Grundlagen der Mathematik". *Abhandlungen aus dem mathematischen Seminar der Univ. Hamburg*, 6:89–92, 1928. English translation *On the Consistency of Arithmetic* in [Heijenoort, 1971a, pp. 485–489].

[Bitbol and Gayon, 2006] Michel Bitbol and Jean Gayon, editors. *L'épistémologie française, 1830–1970*. Presses Universitaires de France, Paris, 2006.

[Brady, 2000] Geraldine Brady. *From Peirce to Skolem: A Neglected Chapter in the History of Logic*. North-Holland (Elsevier), 2000.

[Brouwer, 1925a] Luitzen Brouwer. Zur Begründung der intutionistischen Mathematik I. *Mathematische Annalen*, 93:244–257, 1925.
[Brouwer, 1925b] Luitzen Brouwer. Zur Begründung der intutionistischen Mathematik II. *Mathematische Annalen*, 95:453–472, 1925.
[Brouwer, 1926] Luitzen Brouwer. Zur Begründung der intutionistischen Mathematik III. *Mathematische Annalen*, 96:451–488, 1926.

[Cantor, 1932] Georg Cantor. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Springer, 1932. Ed. by Ernst F. E. Zermelo, reprinted by Georg Olms Verlag, Hildesheim, 1962.
[Cartwright, 1965] Mary L. Cartwright. Jacques Hadamard (1865–1963). *Biographical Memoirs of Fellows of the Royal Society*, 11:75–99, 1965.
[Chang and Lee, 1973] Chin-Liang Chang and Richard Char-Tung Lee. *Symbolic logic and mechanical theorem proving*. Academic Press (Elsevier), 1973.
[Chevalley and Lautman, 1931] Claude Chevalley and Albert Lautman. Notice biographique sur Herbrand. *Annuaire de l'Association amicale de secours des anciens élèves de l'École Normale Supérieure*, 1931:13–15, 1931. Also in [Herbrand, 1968, pp. 13–15]. English translation *Biographical Note on Herbrand* in [Herbrand, 1971, pp. 25–28].
[Chevalley, 1935] Claude Chevalley. Sur la pensée de Jacques Herbrand. *L'Enseignement Mathématique*, 34:97–102, 1935. Also in [Herbrand, 1968, pp. 17–20]. English translation *On Herbrand's Thought* in [Herbrand, 1971, pp. 25–28].
[Chevalley, 1982] Claude Chevalley. Jacques Herbrand. 1982. In [Stern, 1982, pp. 1–2].
[Craig, 1957a] William Craig. Linear reasoning — a new form of the Herbrand–Gentzen Theorem. *J. Symbolic Logic*, 22:250–268, 1957.
[Craig, 1957b] William Craig. Three uses of the Herbrand–Gentzen Theorem in relating model theory and proof theory. *J. Symbolic Logic*, 22:269–285, 1957.

[Dalen, 1974] Dirk van Dalen. Review: Jacques Herbrand: Logical writings by Warren Goldfarb. *The Journal of Philosophy*, 71:544–549, 1974.
[Davis and Putnam, 1960] Martin Davis and Hilary Putnam. A computing procedure for quantification theory. *J. of the ACM*, 7:201–215, 1960.
[Davis, 1963] Martin Davis. Eliminating the irrelevant from mechanical proofs. In *Proceedings of Symposia in Applied Mathematics*, volume 15, pages 15–30, 1963. Also in [Siekmann and Wrightson, 1983, Vol. 1, pp. 315–330].
[Dawson Jr., 1991] John W. Dawson Jr. The reception of Gödel's incompleteness theorems. 1991. In [Drucker, 1991, pp. 84–100].
[Dawson Jr., 1993] John W. Dawson Jr. Prelude to recursion theory: the Gödel–Herbrand correspondence. 1993. In [Wołkowski, 1993, pp. 1–13].
[Delzell, 1996] Charles N. Delzell. Kreisel's unwinding of Artin's proof. 1996. In [Odifreddi, 1996, pp. 113–246].
[Dieudonné, 1982] Jean Dieudonné. Jacques Herbrand et la théorie des nombres. 1982. In [Stern, 1982, pp. 3–7].
[Dieudonné, 1999] Jean Dieudonné. Claude Chevalley. *Transformation Groups*, 4:103–113, 1999. Dedicated to the memory of Claude Chevalley.
[Dreben and Denton, 1963] Burton Dreben and John Denton. A supplement to Herbrand. *J. Symbolic Logic*, 31(3):393–398, 1963.
[Dreben and Denton, 1970] Burton Dreben and John Denton. Herbrand-style consistency proofs. 1970. In [Myhill *et al.*, 1970, pp. 419–433].
[Dreben *et al.*, 1963] Burton Dreben, Peter B. Andrews, and Stål Aanderaa. False lemmas in Herbrand. *Bull. AMS*, 69:699–706, 1963.
[Drucker, 1991] Thomas Drucker, editor. *Perspectives on the History of Mathematical Logic*. Birkhäuser (Springer), 1991.
[Dubreil, 1983] Paul Dubreil. Souvenirs d'un boursier Rockefeller 1929–1931. *Cahiers du séminaire d'histoire des mathématiques*, 4:61–73, 1983.
[Dubucs and Égré, 2006] Jacques Dubucs and Paul Égré. Jacques Herbrand. 2006. In [Bitbol and Gayon, 2006, pp. 301–322].

[Enderton, 1972] Herbert B. Enderton. *A Mathematical Introduction to Logic*. Academic Press (Elsevier), 1972.

[Feferman, 1996] Sol(omon) Feferman. Kreisel's "unwinding" program. 1996. In [Odifreddi, 1996, pp. 247–273].

[Feigenbaum and Feldman, 1963] Edward A. Feigenbaum and J. Feldman, editors. *Computers and Thought*. McGraw-Hill, 1963.

[Frege, 1879] Gottlob Frege. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Verlag von L. Nebert, Halle an der Saale, 1879. pp. III–VIII and pp. 1–54 also in [Berka and Kreiser, 1973, pp. 48–106]. English translation in: [Heijenoort, 1971a], pp. 1–82.

[Frege, 1884] Gottlob Frege. *Die Grundlagen der Arithmetik*. Verlag von M. & H. Marcus, Breslau, 1884. Reprinted 1934, as facsimile: Georg Olms Verlag, Hildesheim, 1990.

[Frege, 1893/1903] Gottlob Frege. *Grundgesetze der Arithmetik – Begriffsschriftlich abgeleitet*. Verlag von Hermann Pohle, Jena, 1893/1903. As facsimile with corrigenda by Christian Thiel: Georg Olms Verlag, Hildesheim, 1998.

[Gentzen, 1935] Gerhard Gentzen. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1935. Also in [Berka and Kreiser, 1973, pp. 192–253]. English translation in [Gentzen, 1969].

[Gentzen, 1936] Gerhard Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, 112:493–565, 1936. English translation in [Gentzen, 1969].

[Gentzen, 1938] Gerhard Gentzen. Die gegenwärtige Lage in der mathematischen Grundlagenforschung – Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften, Neue Folge*, 4:3–44, 1938. Reprinted as facsimile by Wissenschaftliche Buchgesellschaft, Darmstadt, Verlag von S. Hirzel, Stuttgart. English translation in [Gentzen, 1969].

[Gentzen, 1943] Gerhard Gentzen. Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen*, 119:140–161, 1943. Received July 9, 1942. English translation in [Gentzen, 1969].

[Gentzen, 1969] Gerhard Gentzen. *The Collected Papers of Gerhard Gentzen*. North-Holland (Elsevier), 1969. Ed. by Manfred E. Szabo.

[Gilmore, 1960] Paul C. Gilmore. A proof method for quantification theory: Its justification and realization. *IBM Journal Research and Development*, 4:28–35, 1960. Also in [Siekmann and Wrightson, 1983, Vol. I, pp. 151–161].

[Gödel, 1930] Kurt Gödel. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik*, 37:349–360, 1930. With English translation also in [Gödel, 1986ff., Vol. I, pp. 102–123].

[Gödel, 1931] Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. With English translation also in [Gödel, 1986ff., Vol. I, pp. 145–195]. English translation also in [Heijenoort, 1971a, pp. 596–616].

[Gödel, 1934] Kurt Gödel. *On Undecidable Propositions of Formal Mathematical Systems*. The Inst. for Advanced Study, 1934. Lecture notes by Stephen C. Kleene and J. Barkley Rosser. With English translation also in [Gödel, 1986ff., Vol. I, pp. 346–372].

[Gödel, 1939] Kurt Gödel. Consistency proof for the generalized continuum hypothesis. In *Proc. of the National Academy of Sciences, U.S.A., 25*, pages 220–224, 1939. Also in [Gödel, 1986ff., Vol. II, pp. 28–32].

[Gödel, 1986ff.] Kurt Gödel. *Collected Works*. Oxford Univ. Press, 1986ff. Ed. by Sol(omon) Feferman, John W. Dawson Jr., Jean van Heijenoort, Stephen C. Kleene, *et al.*.

[Goldfarb and Scanlon Jr., 1974] Warren Goldfarb and Thomas M. Scanlon Jr. The ω-consistency of number theory via Herbrand's theorem. *J. Symbolic Logic*, 39:678–692, 1974.

[Goldfarb, 1993] Warren Goldfarb. Herbrand's error and Gödel's correction. *Modern Logic*, 3:103–118, 1993.

[Gonseth, 1941] Ferdinand Gonseth, editor. *Les Entretiens de Zurich sur les Fondements et la Méthode des Sciences Mathématiques, Dec. 6–9, 1938*. S. A. Leemann frères & Cie., Zürich, 1941.

[Hadamard, 1949] Jacques S. Hadamard. *The Psychology of Invention in the Mathematical Field*. Princeton Univ. Press, 1949. 2nd rev. ed. (1st ed. 1945).

[Hasse and Herbrand Sr., 1931] Helmut Hasse and Jacques Herbrand Sr.. Letters exchanged between Helmut Hasse and Jacques Herbrand Sr.. Available from the Niedersächsische Staats- und Universitätsbibliothek – Abteilung für Handschriften und seltene Drucke, Göttingen, Germany, 1931.

[Heijenoort, 1968] Jean van Heijenoort. On the relation between the falsifiability tree method and the Herbrand method in quantification theory. Unpublished typescript, Nov. 20, 1968, 12 pp.; Jean van Heijenoort Papers, 1946–1988, Archives of American Mathematics, Center for American History, Univ. of Texas at Austin, Box 3.8/86-33/1. Copy in Anellis Archives, 1968.

[Heijenoort, 1971a] Jean van Heijenoort. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard Univ. Press, 1971. 2nd rev. ed. (1st ed. 1967).

[Heijenoort, 1971b] Jean van Heijenoort. Reviews of articles by G. E. Mints. *J. Symbolic Logic*, 36:524–528, 1971.

[Heijenoort, 1975] Jean van Heijenoort. Herbrand. Unpublished typescript, May 18, 1975, 15 pp.; Jean van Heijenoort Papers, 1946–1988, Archives of American Mathematics, Center for American History, Univ. of Texas at Austin, Box 3.8/86-33/1. Copy in Anellis Archives, 1975.

[Heijenoort, 1982] Jean van Heijenoort. L'œuvre logique de Herbrand et son contexte historique. 1982. In [Stern, 1982, pp. 57–85].

[Heijenoort, 1986a] Jean van Heijenoort. Absolutism and relativism in logic. 1986. In [Heijenoort, 1986b, pp. 75–83].

[Heijenoort, 1986b] Jean van Heijenoort. *Selected Essays*. Bibliopolis, Napoli, copyright 1985. Also published by Librairie Philosophique J. Vrin, Paris, 1986, 1986.

[Heijenoort, 1986c] Jean van Heijenoort. Herbrand's work in logic and its historical context. 1986. In [Heijenoort, 1986b, pp. 99–121].

[Herbrand ..., ...] Please search for labels on Herbrand in the bibliography preceeding this list of references!

[Heyting, 1930a] Arend Heyting. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preußischen Akademie der Wissenschaften, Berlin, Physikalisch-mathematische Klasse II*, 1930:42–56, 1930. Short version also in [Berka and Kreiser, 1973], pp. 173–178.

[Heyting, 1930b] Arend Heyting. Die formalen Regeln der intuitionistischen Mathematik. *Sitzungsberichte der Preußischen Akademie der Wissenschaften, Berlin, Physikalisch-mathematische Klasse II*, 1930:57–71, 158–169, 1930.

[Heyting, 1971] Arend Heyting. *Intuitionism — An Introduction*. North-Holland (Elsevier), 1971. 3rd rev. ed. (1st ed. 1956).

[Hilbert and Ackermann, 1928] David Hilbert and Wilhelm Ackermann. *Grundzüge der theoretischen Logik*. Springer, 1928.

[Hilbert and Bernays, 1968/70] David Hilbert and Paul Bernays. *Die Grundlagen der Mathematik*. Springer, 1968/70. 2nd rev. ed. (1st ed. 1934/39).

[Hilbert, 1922] David Hilbert. Neubegründung der Mathematik (Erste Mitteilung). *Abhandlungen aus dem mathematischen Seminar der Univ. Hamburg*, 1:157–177, 1922. Reprinted with notes in [Hilbert, 1932ff., Vol. 3, pp. 157–177].

[Hilbert, 1926] David Hilbert. Über das Unendliche. *Mathematische Annalen*, 88(151–165), 1926. Received June 24, 1925. English translation *On the infinite* by Stefan Bauer-Mengelberg with an introduction by Jean van Heijenoort in [Heijenoort, 1971a, pp. 367–392].

[Hilbert, 1928] David Hilbert. Die Grundlagen der Mathematik – Vortrag, gehalten auf Einladung des Mathematischen Seminars im Juli 1927 in Hamburg. *Abhandlungen aus dem mathematischen Seminar der Univ. Hamburg*, 6:65–85, 1928. English translation *The foundations of mathematics* by Stefan Bauer-Mengelberg and Dagfinn Føllesdal with a short introduction by Jean van Heijenoort in [Heijenoort, 1971a, pp. 464–479].

[Hilbert, 1931] David Hilbert. Die Grundlegung der elementaren Zahlenlehre. *Mathematische Annalen*, 104(485–494), 1931.

[Hilbert, 1932ff.] David Hilbert. *Gesammelte Abhandlungen*. Springer, 1932ff..

[Hintikka, 1996] K. Jaakko J. Hintikka. *The Principles of Mathematics Revisited*. Cambridge Univ. Press, 1996.

[Howard and Rubin, 1998] Paul Howard and Jean E. Rubin. *Consequences of the Axiom of Choice*. American Math. Society, 1998.

[Kant, 1787] Immanuël Kant. *Critik der reinen Vernunft*. Johann Friedrich Hartknoch, Riga, 1787. 2nd rev. ed. (1st ed. 1781).

[Kleene, 1938] Stephen Cole Kleene. On notation for ordinal numbers. *J. Symbolic Logic*, 3(150–155), 1938.

[Kleene, 1952] Stephen Cole Kleene. *Introduction to Metamathematics*. North-Holland (Elsevier), 1952.

[Kreisel, 1951] Georg Kreisel. On the interpretation of non-finitist proofs — Part I. *J. Symbolic Logic*, 16:241–267, 1951.

[Kreisel, 1952] Georg Kreisel. On the interpretation of non-finitist proofs — Part II. *J. Symbolic Logic*, 17:43–58, iv, 1952.

[Kreisel, 1958] Georg Kreisel. Mathematical significance of consistency proofs. *J. Symbolic Logic*, 23:155–182, 1958.

[Kreisel, 1982] Georg Kreisel. Finiteness theorems in arithmetic: an application of Herbrand's theorem for Σ_2-formulas. 1982. In [Stern, 1982, pp. 39–55].

[Lemmermeyer and Roquette, 2006] Franz Lemmermeyer and Peter Roquette. *Helmut Hasse–Emmy Noether — Die Korrespondenz 1925–1935*. Universitätsverlag Göttingen, 2006.

[Löwenheim, 1915] Leopold Löwenheim. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, 76:228–251, 1915. English translation *On Possibilities in the Calculus of Relatives* by Stefan Bauer-Mengelberg with an introduction by Jean van Heijenoort in [Heijenoort, 1971a, pp. 228–251].

[Loveland, 1978] Donald W. Loveland. *Automated Theorem Proving: A Logical Basis*. North-Holland (Elsevier), 1978.

[Marchisotto and Smith, 2007] Elena Anne Marchisotto and James T. Smith. *The Legacy of Mario Pieri in Geometry and Arithmetic*. Birkhäuser (Springer), 2007.

[Miller, 1992] Dale A. Miller. Unification under a mixed prefix. *J. Symbolic Computation*, 14(321–358), 1992.

[Moore, 1987] Gregory H. Moore. A house divided against itself: The emergence of first-order logic as the basis for mathematics. 1987. In [Phillips, 1987, pp. 98–136].

[Myhill *et al.*, 1970] John Myhill, Akiko Kino, and Richard Vesley, editors. *Intuitionism and Proof Theory. Proceedings of the International Conference Intuitionism and Proof Theory, Buffalo, 1968*. North-Holland (Elsevier), 1970.

[Neumann, 1925] John von Neumann. Eine Axiomatisierung der Mengenlehre. *J. für die reine und angewandte Mathematik (Crellesches J.)*, 154:219–240, 1925. Correction in [Neumann, 1926]. English translation *An Axiomatization of Set Theory* by Stefan Bauer-Mengelberg with an introduction by Jean van Heijenoort in [Heijenoort, 1971a, pp. 393–413].

[Neumann, 1926] John von Neumann. Berichtigung zu der Arbeit „Eine Axiomatisierung der Mengenlehre". *J. für die reine und angewandte Mathematik (Crellesches J.)*, 155:128, 1926. Correction of [Neumann, 1925].

[Neumann, 1927] John von Neumann. Zur Hilbertschen Beweistheorie. *Mathematische Zeitschrift*, 26:1–46, 1927. Received July 29, 1925. *(First occurrence of the notion of* basic typus *(Grundtypus), the cardinal notion in the theory of ε-substitution. Clarification that the consitency proof of [Ackermann, 1925] does not include second order with comprehension.)*

[Neumann, 1928] John von Neumann. Die Axiomatisierung der Mengenlehre. *Mathematische Zeitschrift*, 27(669–752), 1928.

[Newell *et al.*, 1957] Allen Newell, J. C. Shaw, and Herbert A. Simon. Empirical explorations with the Logic Theory Machine: A case study in heuristics. In *Proc. Western Joint Computer Conf.*, pages 218–230. The Institute of Radio Engineers, New York, 1957. Also in [Feigenbaum and Feldman, 1963, pp. 109–133]. Also in [Siekmann and Wrightson, 1983, Vol. I, pp. 49–73].

[Nicod, 1917] Jean Nicod. A reduction in the number of primitive propositions of logic. *Proc. Cambridge Philosophical Soc.*, 19:32–44, 1917.

[Nonnengart, 1996] Andreas Nonnengart. Strong skolemization. Research Report MPI-I-96-2-010, Max-Planck-Institut für Informatik, Im Stadtwald, D–66123 Saarbrücken, 1996.

[Odifreddi, 1996] Piergiorgio Odifreddi, editor. *Kreiseliana — About and Around Georg Kreisel*. A K Peters, Wellesley (MA), 1996.

[Parsons, 1998] Charles Parsons. Finitism and intuitive knowledge. 1998. In [Schirn, 1998, pp. 249–270].

[Peirce, 1885] Charles S. Peirce. On the algebra of logic: A contribution to the philosophy of notation. *American J. of Mathematics*, 7:180–202, 1885. Also in [Peirce, 1993, pp. 162–190].

[Peirce, 1993] Charles S. Peirce. *Writings of Charles S. Peirce — A Chronological Edition, Vol. 5, 1884–1886*. 1993. Ed. by Christian J. W. Kloesel.

[Phillips, 1987] Esther R. Phillips, editor. *Studies in the History of Mathematics*. Mathematics Association of America, Washington (DC), 1987.

[Pieri, 1907/8] Mario Pieri. Sopra gli assiomi aritmetici. *Il Bollettino delle seduta della Accademia Gioenia di Scienze Naturali in Catania*, Series 2, 1–2:26–30, 1907/8. English translation *On the Axioms of Arithmetic* in [Marchisotto and Smith, 2007, § 4.2, pp. 308–313].

[Presburger, 1930] Mojżesz Presburger. Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In *Sprawozdanie z I Kongresu metematyków krajów słowianskich, Warszawa 1929, (Comptes-rendus du 1re Congrès des Mathématiciens des Pays Slaves, Varsovie 1929)*, pages 92–101+395, 1930. Remarks and English translation: [Stanisfer, 1984].

[Reid, 1970] Constance Reid. *Hilbert*. Springer, 1970.

[Ribet, 1976] Kenneth A. Ribet. A modular construction of unramified p-extensions of $\mathbf{Q}(\mu p)$. *Inventiones mathematicae*, 34:151–162, 1976.

[Robinson and Voronkow, 2001] Alan Robinson and Andrei Voronkow, editors. *Handbook of Automated Reasoning*. Elsevier, 2001.

[Robinson, 1965] J. Alan Robinson. A machine-oriented logic based on the resolution principle. *J. of the ACM*, 12:23–41, 1965. Also in [Siekmann and Wrightson, 1983, Vol. 1, pp. 397–415].

[Roquette, 2000] Peter Roquette. On the history of Artins L-functions and conductors — seven letters from Artin to Hasse in the year 1930. *Mitteilungen der Mathematischen Gesellschaft in Hamburg*, 19:5–50, 2000. Web version updated July 23, 2003.

[Roquette, 2007] Peter Roquette. The correspondence Helmut Hasse–Jacques Herbrand. Personal e-mail communications with Peter Roquette. Cf. also [Roquette, 2008], 2007.

[Roquette, 2008] Peter Roquette, editor. *Helmut Hasse–Jacques Herbrand — The complete correspondence, including the six letters of Herbrand to Hasse, dated from May 11 to July 23, 1931, and a letter from Joseph H. M. Wedderburn to Hasse on Herbrand's planned visit to Princeton, dated May 25, 1931*. To appear.

[Rosser, 1936] J. Barkley Rosser. Extensions of some theorems of Gödel and Church. *J. Symbolic Logic*, 1:87–91, 1936.

[Rubin and Rubin, 1985] Herman Rubin and Jean E. Rubin. *Equivalents of the Axiom of Choice*. North-Holland (Elsevier), 1985. 2nd rev. ed. (1st ed. 1963).

[Scanlon Jr., 1973] Thomas M. Scanlon Jr. The consistency of number theory via Herbrand's Theorem. *J. Symbolic Logic*, 38:29–58, 1973.

[Schirn, 1998] Matthias Schirn, editor. *The Philosophy of Mathematics Today*. Oxford Univ. Press, 1998.

[Schröder, 1895] Ernst Schröder. *Vorlesungen über die Algebra der Logik, Vol. 3, Algebra der Logik und der Relative, Vorlesungen I-XII*. B. G. Teubner Verlagsgesellschaft, Leipzig, 1895. English translation of some parts in [Brady, 2000].

[Schütte, 1960] Kurt Schütte. *Beweistheorie*. Springer, 1960.

[Sieg, 2005] Wilfried Sieg. Only two letters: The correspondence between Herbrand and Gödel. *The Bulletin of Symbolic Logic*, 11(2):172–184, 2005.

[Siekmann and Wrightson, 1983] Jörg Siekmann and Graham Wrightson, editors. *Automation of Reasoning*. Springer, 1983.

[Siekmann, 1989] Jörg Siekmann. Unification theory. *J. Symbolic Logic*, 7:207–274, 1989.

[Skolem, 1919] Thoralf Skolem. Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen. *Skrifter, Norske Videnskaps-Akademi i Oslo (= Videnskapsselskapet i Kristiania), Matematisk-Naturvidenskapelig Klasse, J. Dybwad, Oslo*, 1919/3:1–37, 1919. Dated May 26, 1917. Also in [Skolem, 1970, pp. 67–101]. *(First explicit occurrence of quantifier elimination).*

[Skolem, 1920] Thoralf Skolem. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit und Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen. *Skrifter, Norske Videnskaps-Akademi i Oslo (= Videnskapsselskapet i Kristiania), Matematisk-Naturvidenskapelig Klasse, J. Dybwad, Oslo*, 1920/4:1–36, 1920. Also in [Skolem, 1970, pp. 103–136]. English translation of § 1 *Logico-Combinatorial Investigations in the Satisfiability or Provability of Mathematical Propositions: A simplified proof of a theorem by Leopold Löwenheim and generalizations of the theorem* by Stefan Bauer-Mengelberg with an introduction by Jean van Heijenoort in [Heijenoort, 1971a, pp. 252–263]. *(Löwenheim–Skolem Theorem via Skolem normal form and choice of a sub-model).*

[Skolem, 1923a] Thoralf Skolem. Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich. *Skrifter, Norske Videnskaps-Akademi i Oslo (= Videnskapsselskapet i Kristiania), Matematisk-Naturvidenskapelig Klasse, J. Dybwad, Oslo*, 1923/6:1–38, 1923. Written 1919. Also in [Skolem, 1970, pp. 153–188]. English translation *The foundations of elementary arithmetic established by means of the recursive mode of thought, without the use of apparent variables ranging over infinite domains* by Stefan Bauer-Mengelberg with an introduction by Jean van Heijenoort in [Heijenoort, 1971a, pp. 302–333].

[Skolem, 1923b] Thoralf Skolem. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Proc. 5th Scandinaviska Matematikerkongressen, Helsingfors, July 4–7, 1922*, pages 217–232, Helsingfors, 1923. Akademiska Bokhandeln. Also in [Skolem, 1970, pp. 137–152]. English translation *Some remarks on Axiomatized Set Theory* by Stefan Bauer-Mengelberg with an introduction by Jean van Heijenoort in [Heijenoort, 1971a, pp. 290–301]. *(The best written and, together with [Skolem, 1920], the most relevant publication on the Löwenheim–Skolem Theorem; although there is some minor gap in the proof according to Wang [1970], cured in [Skolem, 1929]. Includes Skolem's Paradox and a proof of the Löwenheim–Skolem Theorem which does not require any weak forms of the Axiom of Choice, via Skolem normal form and construction of a model without assuming the previous existence of another one).*

[Skolem, 1928] Thoralf Skolem. Über die mathematische Logik (Nach einem Vortrag gehalten im Norwegischen Mathematischen Verein am 22. Oktober 1928). *Nordisk Matematisk Tidskrift*, 10:125–142, 1928. Also in [Skolem, 1970, pp. 189–206]. English translation *On Mathematical Logic* by Stefan Bauer-Mengelberg and Dagfinn Føllesdal with an introduction by Burton Dreben and Jean van Heijenoort in [Heijenoort, 1971a, pp. 508–524]. *(First explicit occurrence of Skolemization and Skolem functions).*

[Skolem, 1929] Thoralf Skolem. Über einige Grundlagenfragen der Mathematik. *Skrifter, Norske Videnskaps-Akademi i Oslo (= Videnskapsselskapet i Kristiania), Matematisk-Naturvidenskapelig Klasse, J. Dybwad, Oslo*, 1929/4:1–49, 1929. Also in [Skolem, 1970, pp. 227–273]. *(Detailed discussion of Skolem's Paradox and the Löwenheim–Skolem Theorem).*

[Skolem, 1941] Thoralf Skolem. Sur la portée du théorème de Löwenheim–Skolem. 1941. In [Gonseth, 1941, pp. 25–52]. Also in [Skolem, 1970, pp. 455–482].

[Skolem, 1970] Thoralf Skolem. *Selected Works in Logic*. Universitetsforlaget Oslo, 1970. Ed. by Jens E. Fenstad. *(Without index, but with most funny spellings in the newly set titles).*

[Smullyan, 1968] Raymond M. Smullyan. *First-Order Logic*. Springer, 1968.

[Stanisfer, 1984] Ryan Stanisfer. *Presburger's Article on Integer Arithmetic: Remarks and Translation*. Dept. of Computer Science, Cornell Univ., Ithaca, NY, 1984.

[Stern, 1982] Jacques Stern, editor. *Proc. of the Herbrand Symposium, Logic Colloquium'81, Marseilles, France, July 1981*. North-Holland (Elsevier), 1982.

[Tait, 1981] William W. Tait. Finitism. *J. of Philosophy*, 78:524–546, 1981.

[Tait, 2006] William W. Tait. Gödel's correspondence on proof theory and constructive mathematics. *Philosophia Mathematica (III)*, 14:76–111, 2006.

[Tarski, 1951] Alfred Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, Berkeley and Los Angeles, 1951. 2nd rev. ed. (1st ed.: Rand Corporation, Santa Monica (CA), 1948).

[Vihan, 1995] Přemysl Vihan. The last months of Gerhard Gentzen in Prague. *Collegium Logicum — Annals of the Kurt Gödel Society*, 1:1–7, 1995.

[Wallen, 1990] Lincoln A. Wallen. *Automated Proof Search in Non-Classical Logics*. MIT Press, 1990.

[Wang, 1970] Hao Wang. A survey of Skolem's work in logic. *In [Skolem, 1970]*, pages 17–52, 1970.

[Weyl, 1921] Hermann Weyl. Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift*, 14:39–79, 1921.

[Weyl, 1928] Hermann Weyl. Diskussionsbemerkungen zu dem zweiten Hilbertschen Vortrag über die Grundlagen der Mathematik. *Abhandlungen aus dem mathematischen Seminar der Univ. Hamburg*, 6:86–88, 1928.

[Whitehead and Russell, 1910–1913] Alfred North Whitehead and Bertrand Russell. *Principia Mathematica*. Cambridge Univ. Press, 1910–1913. 1st ed. (2nd rev. ed. 1925ff.).

[Wirth, 2004] Claus-Peter Wirth. Descente Infinie + Deduction. *Logic J. of the IGPL*, 12:1–96, 2004.

[Wirth, 2006] Claus-Peter Wirth. lim+, δ^+, *and Non-Permutability of* β*-Steps*. SEKI-Report SR–2005–01 (ISSN 1437–4447). SEKI Publications, Saarland Univ., 2006. rev. ed. of July 30, 2006. `http://www.ags.uni-sb.de/~cp/p/nonpermut`.

[Wirth, 2008a] Claus-Peter Wirth. Hilbert's epsilon as an operator of indefinite committed choice. *J. Applied Logic*, 2008. `http://dx.doi.org/10.1016/j.jal.2007.07.009`.

[Wirth, 2008b] Claus-Peter Wirth. Shallow confluence of conditional term rewriting systems. *J. Symbolic Computation*. To appear. `http://www.ags.uni-sb.de/~cp/p/shallow`.

[Wittgenstein, 1939] Ludwig Wittgenstein. *Lectures on the Foundations of Mathematics, Cambridge*. Univ. of Chicago Press, copyright 1989, 1939. Cora Diamond (ed.), from the notes of R. G. Bosanquet, Norman Malcolm, Rush Rhees, and Yorick Smythies.

[Wołkowski, 1993] Zbigniew Wołkowski, editor. *1st Int. Symposium on Gödel's Theorem, Paris, May 27–29, 1991*. World Scientific Publishing, 1993.

[Zach, 2001] Richard Zach. *Hilbert's Finitism: Historical, Philosophical, and Metamathematical Perspectives*. PhD thesis, Univ. of California, Berkeley, 2001.

[Zaremba, 1926] Stanisław Zaremba. *La Logique des Mathématiques*. Mémorial des Sciences Mathématiques, Fascicule XV. Gauthier-Villars, Paris, 1926.

[Zygmunt, 1991] Jan Zygmunt. Mojżesz Presburger: Life and work. *History and Philosophy of Logic*, 12:211–223, 1991.

THE LOGIC OF THE *TRACTATUS*

Michael Potter

Most of the attention that has been paid to the *Tractatus* since it was published in the early 1920s has focused on the metaphysical picture it paints of the world and our role in it. Indeed, one recently influential school interprets the book as an extended *reductio* of the attempt to paint any such picture at all. Whether or not this is so, what is plain is that the book's pretensions extend far beyond matters of pure logic.

Nevertheless, it *is* a treatise in logic as well as a treatise in metaphysics — a 'logico-philosophical treatise', as its full title describes it — and in this handbook we shall concentrate on the account it gives of logic. Notice, though, that I say 'as well', and for this reason it is not possible to give proper consideration to the logical aspects of the work without considering the metaphysical ones too: the method of the *Tractatus* is neither to deduce the structure of our language from that of the world of which it attempts to speak nor, conversely, to deduce the structure of the world from that of language, but rather to deduce the structure of both from the necessary representing relationship between them.

Wittgenstein composed the *Tractatus* while serving in the Austrian army during the First World War. It is possible to trace its genesis through the notebooks in which he worked out many of its ideas between 1914 and the beginning of 1917, and through the so-called *Prototractatus*, the first draft of the work thought to have been composed between 1915 and 1917. Two earlier sets of notes, dictated to Russell in 1913 and Moore in 1914, also help to indicate the order of discovery, but their usefulness as exegetical tools in interpreting the *Tractatus* itself is limited by the fact that some of Wittgenstein's logical views changed somewhat in the meantime.

The principal influences on the logic of the *Tractatus* were Frege, with whom Wittgenstein discussed logical matters on a handful of occasions before the war and corresponded fitfully (although not philosophically) during it; and Russell, who supervised Wittgenstein's work on logic in Cambridge between 1911 and 1913. It is as well to bear these influences in mind throughout one's reading of the work: Wittgenstein's method is frequently to explain his own view only by opposition to Frege's or Russell's.

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

1 OBJECTS

Wittgenstein's dependence on the previous work of Frege and Russell is evident straightaway in what he says about the twin problems of how singular terms refer to objects and how propositions express things about these objects that can be either true or false. By drawing elements from each of their accounts Wittgenstein aimed to avoid the difficulties of both.

The solution he came to involves a conception of the world as consisting of objects standing in various combinations with one another: the ways objects *may* stand in combination are called *states of affairs*. The sum total of the ways they *do* stand is the world. And so a complete description of the world would amount to no more than a way of specifying which states of affairs occur and which do not.

The picture the *Tractatus* paints of the world is thus extremely austere. Its simplicity is deceptive, however: a very large part of the book is taken up with working out the consequences of this picture. Here, though, we shall concentrate on the narrower task of explaining what the view amounts to and why Wittgenstein was led to adopt it.

The first point on which Wittgenstein had to adjudicate between Frege and Russell was whether in the explanation of how singular terms refer there has always to be a mediating element — what Frege called the 'sense' of the term — serving to present in some particular fashion the object referred to, or whether there is at least a subclass of the singular terms — what Russell called 'logically proper names' — which refer to their objects directly without any intermediary.

Frege had adopted his 'two-step' semantics for singular terms in order to explain the difference between 'Hesperus is Phosphorus', which is informative, and 'Hesperus is Hesperus', which is not. The difference could not simply be a matter of the ideas which I associate with the words 'Hesperus' and 'Phosphorus', since these are private to me. Frege believed that when I utter a sentence, what I express — he called it the 'thought' — occupies an inter-subjectively accessible 'third realm' of entities that are neither physical nor mental. What he sought was an account which made the thought expressed different in the two cases. He thought he could achieve this only by associating with the two singular terms different senses, i.e. items lying in this same third realm that explain why the thoughts expressed by the two sentences are different.

Russell, on the other hand, while he shared Frege's view that what the two sentences express are different, thought from 1905 onwards that this difference could be explained by treating 'Hesperus' and 'Phosphorus' not as logically proper names but as disguised definite descriptions. The difference in what is expressed would then emerge through a difference in logical analysis.

But what was Russell's reason for rejecting Frege's two-step view? In 'On denoting' itself the reason he gives is the famous 'Gray's Elegy argument' (so called because one of the central examples he uses to explain the flaw he sees in the two-step view involves the singular term 'The first line of Gray's Elegy'). But

in truth Russell seems to have designed the Gray's Elegy argument to wield not against Frege but against mixed theories such as the one he himself had advanced in the *Principles of Mathematics*. According to that theory *some* singular terms have both what he called a meaning and a denotation (or, in Fregean terms, a sense and a reference), but there remain also logically proper names, which have only a denotation. Although the Gray's Elegy argument seems effective against this target, it is far from clear that it also works against a theory such as Frege's which makes *every* instance of reference go by way of an intermediate sense.

What seems most plausible as an explanation of this lacuna in Russell's argument is that the one-step view was by the time he wrote 'On denoting' too ingrained in Russell's conception to require any argument. It had been one of the cornerstones of his revolt against idealism in 1898 to suppose that propositions succeed in being *about* the world only by containing parts of that very world. At that time he had not yet come across Frege's two-step semantic theory, but when he did so a couple of years later, it would have seemed to him to fall into the old idealism by making sentences be about the third realm of senses rather than about the world itself.

Wittgenstein sided decisively with Russell's one-step view, describing Russell's theory of descriptions as 'quite CERTAINLY correct'. [Wittgenstein, 1995, p61] And Wittgenstein's reason was something like the one we have just ascribed to Russell, namely that Frege's theory does not correctly capture how sentences are expressive of the way the world is. However, Wittgenstein's formulation of his reason is not easy to understand:

> If the world had no substance, then whether a proposition had sense would depend on whether another proposition was true. In that case we could not sketch any picture of the world (true or false). [Wittgenstein, 1961, 2.0211–2.0212]

Not only is this passage highly compressed, but it makes use of notions — substance, sense — which have not yet been defined and, worse, whose precise definitions depend on the very argument we are currently trying to understand. Nevertheless, this is one of the central arguments of the *Tractatus*, and there is little in the book that does not depend in some way on its conclusion, so we need to do what we can to break into its argumentative circle. The best way to do this, I think, is to adopt at this stage only a schematic understanding of the notions of substance and sense. Let us assume at this stage, therefore, no more than that our propositions express something about how the world is or could be: what a propositions expresses is called its *sense*, and *substance* consists of the simple objects whose existence Wittgenstein's argument is intended to prove.

The basis of the argument for substance is Wittgenstein's claim that the sense of one proposition — what it expresses — cannot always be dependent on the truth of another proposition. For if it were, he says, 'it would then be impossible to form a picture of the world (true or false).' Wittgenstein seems to be appealing here to a regress argument: in order that sense should be determinate, it should

not, ultimately, depend on how the world is.

But we are in some danger at this point of returning to the explanatory circle we have only just exited from. For it is not obvious why the regress *should* be thought vicious unless we give the notion of sense the specific technical meaning it has in the *Tractatus*. There the sense of a proposition consists in a division of all the ways the world could be into two classes in such a way that the proposition is true if the way the world is lies in one of the classes, false if it lies in the other. This notion of sense would plainly be incoherent in the case of a proposition whose sense depended on the truth of another proposition, since that other proposition could be false (even if in fact it is not), and in that case the first proposition would not have a sense at all. So its sense would not divide *all* worlds into two classes as required. But the notion of sense in the *Tractatus* has, as we have said, particular features which belong to the same circle of ideas as the notion of substance he is here trying to justify. Shorn of these features, it is not immediately clear why sense should not be determinate *despite* depending on how the world is.

If, however, we grant Wittgenstein his claim that whether propositions have sense must be independent of how the world is, how does it follow that the world has substance — is made up, that is to say, out of simple objects? Elizabeth Anscombe [Anscombe, 1967, ch4] has suggested an argument roughly as follows. Consider the sentence 'A is tall'. There are two sorts of circumstance that would make this sentence false: it might be that there is no person A, tall or not; or it might be, on the other hand, that there is an A, but he is not tall. In this way, the sentence is capable of two sorts of falsity: we may distinguish these by calling them, respectively, external and internal falsity. Is it possible that all propositions are capable of both kinds of falsity? Wittgenstein's view is that it is not. His reason is that in order to understand what it would be for 'A is tall' to have the external kind of falsity, we should have to know what it would be for 'A exists' to be false. The latter sentence says that the constituents of A are arranged in the right way; so in order to grasp its sense we should have to know the truth of a certain proposition describing the construction of A from his constituents and saying that they are configured in such a way as to form A. Plainly, if every sentence were capable of external falsity, we should have to know the truth of some proposition in order to know the sense of another [Wittgenstein, 1961, 2.0211]. But for Wittgenstein that would, as already noted, be absurd. He inferred that there must be propositions that admit of only one kind of negation.

Now Russell's theory provides a way of analysing the distinction between internal and external negation. A sentence will have both sorts of negation just in case its correct logical analysis contains a definite description. If, in our present example, 'A' is really an abbreviation for 'John's only son', Russell's analysis allows us to distinguish the external negation, 'it is not the case that John has just one son who is tall', from the internal one, 'John has just one son and he is not tall'. So if there are, in the final logical analysis, to be sentences with only one negation, and therefore only one sort of truth or falsity, the singular terms they contain must be not disguised definite descriptions but logically proper names. Thus if we accept

Russell's theory of descriptions, we do indeed have an argument that determinacy of sense requires there to be logically proper names.

Notice, though, that this argument, like Russell's in 'On denoting', still does not seem to give us a reason to adopt the distinction between definite descriptions and logically proper names in the first place. For *that* we need to return to the idea that what a proposition expresses (i.e. its sense) should determine a division of ways the world could be (more briefly, *possible worlds*) into those that make the proposition true and those that make it false. Wittgenstein's idea seems to have been that it is incoherent to conceive of some aspects of the world as varying without correspondingly conceiving of other aspects as remaining fixed. If *everything* varies, that is to say, we lose our grip on the very idea of variation. Our notion of variability is inherently relative: something varies in relation to something else; it does not simply vary on its own.

In order for sense to be determinate, then, there need to be fixed elements, hinges, so to speak, around which the possibilities turn. These fixed elements are what Wittgenstein calls *objects*. They are, that is to say, the substance of the world. What will vary between worlds, on this conception, will not be the objects, which are by definition constant between different possible worlds, but the ways in which the objects are arranged. The word Wittgenstein uses for a way objects could be arranged is *Sachverhalt*, which is rendered 'atomic fact' in the first English translation of the *Tractatus* [Wittgenstein, 1922], 'state of affairs' in the second [Wittgenstein, 1961]: here we shall conform with the second and translate *Sachverhalt* as state of affairs.

Now I cautioned at the outset against reading the argument of the *Tractatus* as being conducted either within logic or within metaphysics exclusively, and here we have an instance of this duality being played out. For although it might seem at first as if the argument we have just given for conceiving of the world as made up out of simple objects is purely metaphysical, closer inspection reveals that it is not. The world in question here is a world represented in language. If we dropped that assumption, we could conceive of the world as varying continuously like a fluid: that would allow us to retain our conception of variation as relative but not commit us to simple objects around which the variation is to hinge. Wittgenstein's atomism is therefore essentially dependent on his conception of language as lying in a harmonious representing relation to the world.

Notice, too, how by treating the notion of an object as simply a building block in his construction of the notion of possibility Wittgenstein makes it trivially the case that any two possible worlds have the same objects. If something might not have existed, what that reveals, for Wittgenstein, is that the thing in question is actually complex, since for something not to exist is itself a sort of complexity. And if there were two ways a world could be with no objects at all in common, it would not be correct to describe them as ways the *same* world could be; but rather they would be wholly distinct worlds with nothing whatever to do with one another.

Our argument for simple objects, then, is linguistic as well as metaphysical.

And its conclusion, too, is played out in language as well as in the world. For the way in which simple objects are represented in language is by the simple items which Russell had called logically proper names but which Wittgenstein calls just *names*. And the way in which states of affairs are represented is by concatenations of names called *elementary propositions*.

2 STATES OF AFFAIRS

2.1 Picturing

The fundamental unit in language, we have just noted, is a concatenation of names, which is supposed to correspond to a state of affairs in the world. But *how* does a concatenation of names correspond to a state of affairs? We have seen that Wittgenstein had a choice between Frege's and Russell's theories of how a name refers to an object, and chose Russell's. At this point he once more had a choice, this time between their accounts of the structure of propositions, but now he chose something much more like Frege's account.

Russell's theory had been that a proposition is a sort of complex entity composed of the things it is about. But just as Russell's old theory of denoting faced a difficulty over explaining the connexion between a denoting complex and its denotation, so his old conception of a proposition as a complex faces a difficulty when it comes to explaining the connexion between this complex and the putative fact it expresses. The complex cannot express a fact on its own, so we are naturally tempted to posit a third thing — propositionally expressive glue, so to speak — which combines the elements in the complex in such a way as to make them expressive of the fact. But this is only to postpone the problem, not to solve it, since we are now owed a corresponding explanation of the expressiveness of a new proposition, namely the one which says that the original proposition asserts a fact.

This is known as 'Bradley's regress', and versions of it dogged Russell's attempts to explain the expressive unity of the proposition until he adopted a theory (the multiple relation theory of judgment) which abandoned the idea that the proposition is an entity of any kind at all. Now the multiple relation theory is rejected in the *Tractatus*, but we shall come to Wittgenstein's reasons for this later. In the meantime let us merely contrast Russell with Frege, who avoided the difficulty of Bradley's regress by supposing that what makes a proposition expressive is the unsaturatedness of the predicate or relation symbol it contains. Thus he analysed the sentence 'aRb', for example, as formed by substituting the names 'a' and 'b' into the argument places in the expression 'xRy'. Writing the relational expression as 'xRy' rather than just 'R' is a way of reminding us that it is unsaturated, requiring the addition of two names if it is to form a complete expression (a sentence). Frege does not posit a further element, the 'copula' of traditional logic, gluing 'a', 'b' and 'R' together to form 'aRb', because the relational expression 'xRy' already contains, through its unsaturatedness, the requisite glue.

But how does this explain the expressiveness of 'aRb'? Frege's view was that the unsaturated relation symbol 'xRy' has corresponding to it something which, although not an object, is nevertheless objective: he called it a *relation*. The relation is unsaturated in *just the same* way that the relation symbol is. So if it is determinate what relation 'xRy' expresses and what objects 'a' and 'b' refer to, no further resources are needed to determine the thought expressed by the sentence 'aRb'.

What Wittgenstein's picture theory does is to extract this insight from its Fregean context and transpose it to a Tractarian key. What is lost in the transposition is Frege's division of the objective items into two types, unsaturated ones called concepts and relations, and saturated ones called objects. Wittgenstein saw that it is arbitrary to regard the concept as unsaturated and the object as saturated, and hence abolished this way of distinguishing between them. When it comes to considering the range of possibilities that are expressible in language, simple concepts and relations are just as much constant elements around which the possibilities hinge as Fregean objects are: both, therefore, have equal title to be regarded as objects in the sense of the *Tractatus*.

It is important to realize, though, that what Wittgenstein was doing was *not* to abolish the distinction between objects and concepts, but only to deny that it is a binary distinction based on saturatedness. It remains as important for Wittgenstein as it was for Frege to recognize that the predicate 'xRy' has a different form from the name 'a'. But for Wittgenstein that difference is to be expressed not in terms of saturatedness but rather in terms of the possibilities that exist for combining these expressions with others to form propositions. Thus a Fregean language containing relation symbols and names can be straightforwardly redescribed as a Wittgensteinian language in which the Fregean names are now Tractarian names of a certain form and the Fregean relation symbols are Tractarian names of another form. What Wittgenstein's account leaves open that Frege's does not, however, is that there might be further forms of object that cannot be described in Fregean terms.

2.2 States of affairs

Objects are permanent; what varies between possible worlds is not them but how they are configured to form states of affairs. These configurations are represented linguistically by names concatenated to form elementary propositions. And the objects are configured in the state of affairs in just the same way as the names are in the elementary proposition. What can we deduce from this conception about the structure of the states of affairs?

The first thing to note is that objects need not be substitutable for one another: if a occurs in a state of affairs, there need not be another state of affairs obtainable by substituting b for a in this state of affairs. The range of possibilities of occurrence of a in states of affairs is its *form*. So what we have just noted is that there may be more than one form of object.

Since objects are unchanging, the number of objects of any one form is fixed. But the system does not tell us what that number is: nothing prevents there being only one object of a certain form, for example. The forms themselves, on the other hand, are rather more heavily constrained. Wittgenstein first rules out a form which involved no possibilities of combination at all, not because it is absurd but because it is pointless. An object of such a form would not belong to any states of affairs at all, even non-occurring ones, and the corresponding name would not belong to any elementary propositions, true or false. So there would be no conceivable way of using the name, and hence the name would not really be meaningful. 'If a sign is *useless*, it is meaningless. That is the point of Occam's maxim.' [Wittgenstein, 1961, 3.328].

In the last section we mentioned Wittgenstein's use of the picturing metaphor to explain the way in which an elementary proposition succeeds in expressing a state of affairs. Some commentators have been misled by this metaphor into supposing that the way in which objects combine is to be thought of as itself a substantive item capable of varying even though the objects remain fixed. They have supposed, for instance, that there might be various ways of configuring the same objects a and b to form a state of affairs, ways which we might represent linguistically by putting the name 'a' in various spatial relationships — to the left, above, diagonally to the right — to the name 'b':

$$ab \qquad \begin{matrix} a \\ b \end{matrix} \qquad a^b.$$

Now we can see straightaway that there is no symbolic necessity to represent these states of affairs in this way. We could more conventionally (and with rather greater typographical convenience) represent them, for instance,

$$aLb \quad aAb \quad aEb,$$

using the relation symbols 'xLy', 'xAy' and 'xEy' to play the role of the spatial relationships — to the left, above, diagonally to the right — in the previous notation. In this new notation the states of affairs are now represented as having a third dimension of variability, and the various configurations of a and b have apparently been transformed into objects lying in this third position.

So far, though, we have only produced another means of representation: we have not yet shown that the previous means could not be for Wittgenstein an equally acceptable alternative. To show that it could not, we have to see how it contravenes Wittgenstein's conception of how the symbolism represents.

> If I know an object I also know all its possible occurrences in states of affairs. (Every one of these possibilities must be part of the nature of the object.) A new possibility cannot be discovered later. [Wittgenstein, 1961, 2.0123]

The spatial configurations we considered earlier contravene this constraint, since they allow that I might know a and b as occurring in the state of affairs ab and

yet be wholly unaware of the alternative possibility that they combine to form the state of affairs a^b.

But if 2.0123 shows *that* Wittgenstein wished his symbolism to rule out these configurations, it does not really explain why. This is because his reason is essentially symbolic in origin. If we think of the form of a name as playing the role in Tractarian symbolism that a part of speech plays in ordinary language, then what we are rejecting as incoherent can be thought of as a language in which the same word functions as two different parts of speech. Wittgenstein's point is that any case we are tempted to describe in this way should rather be described as a case of two different words with only a verbal similarity connecting them. You cannot grasp a word without grasping which part of speech it is; and equally once you *have* grasped a word as a particular part of speech, you cannot subsequently discover that the very *same* word also functions as a different part of speech.

It is as important to be clear about what this does not show as about what it does. It does not show that it would be in some way incorrect to represent spatial relations between objects by means of corresponding spatial relations between names. Quite the reverse: as Wittgenstein's choice of the metaphor of picturing strongly suggests, such a means of representation would be for him a paradigm case of representation. The mistake is only to treat the different spatial relationships logically as different grammatical forms rather than as different objects capable of participating in one form.

We can also show by similar means that a single object on its own cannot constitute a state of affairs: every proposition is composite [Wittgenstein, 1961, 5.5261]. For the difficulty with supposing that a single object a occurs as a state of affairs is that our symbolism bars us from conceiving of this state of affairs as *not* obtaining: the object cannot fail to exist, and hence neither can the state of affairs.

> But how CAN a SINGLE *word* be true or false? At any rate it cannot express the *thought* that agrees or does not agree with reality. That *must* be articulated.
>
> A single word cannot be true or false in *this* sense: it cannot agree with reality, or the opposite. [Wittgenstein, 1979, 5 Oct. 1914]

There is, in other words, no such thing as an unarticulated state of affairs. The situation in which all there seems to be is a single object is not correctly representable as such. It ought instead to be described by positing another object — a logical spirit level, if you like — against which a either stands or does not stand.

2.3 Logical and epistemological atomism

According to Wittgenstein, then, we inevitably represent the world as made up from simple entities which he calls objects and which are represented in language by names. He may thus be described as an advocate of *atomism*. So too was Russell, at least during the period that concerns us here. And it was of the

essence of the conception they shared that the link between a name and its object, once established, should be unbreakable. My representation of the world cannot allow for the possibility that a name I use should fail to refer to an object, since to allow for such a possibility is to represent the object as complex and hence to deny that it is what I was representing it as. Where Russell and Wittgenstein differed, and differed radically, was in the conclusions they drew from this need for an unbreakable link between name and object.

Russell interpreted the constraint epistemologically as the requirement that an object I can name should be one about whose existence and identity I cannot be mistaken, and hence one with which I have immediate acquaintance. His paradigm examples of such objects were sense data. Although I may be mistaken about whether I see a tree outside my window — it might, I suppose, be a trick of the light — it is incontrovertible that I am experiencing a tree-like sense datum. But the difficulty with this is that however unmistakable the sense datum, I seem to be able to represent the possibility that it might not have existed.

And yet if I treat sense data as *logically* simple, I cannot really represent such a possibility. I am forced to conclude that the possibility I appear to be able to represent is illusory. All I can truly represent is the world of my experience, and talk of other different experiences I could have had but didn't is really a misleading way of talking about the experiences I have actually had.

For Wittgenstein, by contrast, the argument moves in the opposite direction. He takes the very notion of representation to involve that of possibility, since to represent the world as being a certain way is to represent it as being that way as opposed to other ways it could have been. Since nothing can be both contingent and simple, it follows in particular that our experience, which we represent as contingent, is not simple. To interpret objects as sense data would therefore be to restrict severely the applicability of the *Tractatus*. Tractarian objects cannot be anything we normally represent as items of experience, since we represent all such things as contingent. This is not to say, of course, that a sense-datum language — a language in which the names refer to sense data — would be unTractarian, but merely that our representational practices, which represent sense data as contingent, give us reason to deny that *our* language is a sense-datum language.

Russell coined the phrase 'logical atomism' in 1911 to describe his own position, but a brief way to put the contrast might be to describe Russell's atomism rather as *epistemological*, and reserve the term *logical atomism* for Wittgenstein's. It is a contrast that runs deep in their philosophies of the period and is implicated in almost every other difference between them, but a failure to appreciate it has nevertheless been one of the most common misreadings of the *Tractatus*. The reason for this misreading cannot be traced to the influence of Russell's atomistic writings alone, however. It is due also to the espousal of the *Tractatus* by the logical positivists of the Vienna Circle, whose scientific empiricism made it hard for them to read the work any other way.

The tendency to read the *Tractatus* as committed to an identification of objects with sense data is exacerbated by the fact that Wittgenstein himself had some

sympathy with this idea both before he wrote it and afterwards. Thus in the *Notebooks* as late as 1915 we find him identifying the concept of the object with the concept 'this'; and two days later he grants that it seems

> perfectly possible that patches in our visual field are simple objects . . . ; the visual appearances of stars even seem certainly to be so. [Wittgenstein, 1979, 18 June 1915]

And when he admits a decade after the *Tractatus* that 'I do not now have phenomenological language, or "primary language" as I used to call it, in mind as my goal', [Wittgenstein, 1975, I, sn1] it is hard not to interpret him as implying that he at one time believed the correct analysis of language to be phenomenological.

But whatever his views before and after he wrote the *Tractatus*, in the book itself Wittgenstein remains non-committal as to the identity of the objects. He goes as far as to say that space, time and colour are forms of objects, but it is a little surprising to find him saying even this, since it is an isolated remark that does not lead anywhere. And aside from this one remark he simply ignores the question as being outside the remit of the logician.

What we have seen, then, is that Wittgenstein is not concerned in the *Tractatus* with studying parochial features of the epistemology of the language we happen to find ourselves using. That is not to say, though, that it is a work in which epistemology does not feature. Indeed where epistemology occurs it can be located quite precisely. Among the possible worlds which our language allows for, there is one which is privileged by virtue of being the actual world. But this privilege cannot be a feature of the representational scheme itself: once the scheme has been fixed, there is no *logical* difference between the actual world and any of the others. What distinguishes the actual world, therefore, is only that it is the one we experience. We shall see later how this distinction impacts on Wittgenstein's account of the self.

2.4 Realism in the Tractatus

The *Philosophical Investigations*, largely composed during the Second World War, was the first of Wittgenstein's later (post-1930) works to be published, and it encouraged the notion that his later position was opposed to, and represented a clean break from, that of the *Tractatus*. A feature of this characterisation was the attribution to the *Tractatus* of a certain kind of realism, namely the view that the facts of the world are wholly independent of the subject. That this misconstrues the *Tractatus* is clear from Wittgenstein's equation of realism with solipsism [Wittgenstein, 1961, 5.64]. His realism, if that is what it is, was intended to avoid the perspective from which the self and the world are separate entities. Instead, and solipsistically, he regarded the self, not as a part of the world, but as constituting its limits. This is plainly not realism in the traditional sense in which it is opposed to idealism. Indeed Wittgenstein once remarked that realism, idealism and solipsism coincide. The reason is that once we have adopted the

conception Wittgenstein is recommending, according to which the self (i.e. the representing relation between language and the world) shrinks to an extensionless point, all that is left is language or, equally, the world. We were tempted at first to see the gap between these two — the gap on either side of which realism and idealism stand — as being bridged by the self, which represents one by means of the other. But now the gap has simply disappeared.

Perhaps, though, Wittgenstein's realism was of another sort: perhaps it says something, not about the relation between the subject and the world, but about that between our representations and the world they are supposed to represent. Perhaps it says that the world is independent of our attempts to depict it in language. We can assess this proposal by considering the Tractarian doctrine of internal properties. The internal properties of an object are those it could not but have: it is unthinkable that this object should not possess those properties [Wittgenstein, 1961, 4.123, 2.012]. Its external properties, on the other hand, are given by those configurations into which it actually does enter [Wittgenstein, 1961, 2.0123-2.01231].

Now the form of an object consists in its possibilities of occurrences in states of affairs, i.e. the ways in which it can hang together with other objects to form a state of affairs. But it might be that two objects had the same form. It would then be the case that they possessed certain essential features that had no bearing on their possibilities of occurrence in states of affairs. In that case, such features would fail to find expression within language in the syntactic rules governing names standing for the objects, since these are determined only by the possible states of affairs into which the corresponding objects can enter. So if there are two objects with the same possibilities of configuration but divergent essential properties, there will be nothing in the signs themselves or in their possibilities of combination in virtue of which, of two names 'a' and 'b', one refers to one of these objects and one to the other. It cannot be said, because we can only say things about external properties. But could it be shown? After all, there would be nothing in the logical rules of language to show it: for any context $\_\_x\_\_$, '$\_\_a\_\_$' would make sense just in case '$\_\_b\_\_$' made sense. The answer must presumably be that 'a' is shown to mean a rather than b by our *use* of 'a' to mean a.

Notice, though, that what is involved in showing this need not be especially mysterious. How it comes about is by means of what Wittgenstein calls 'elucidations'. These constitute the one point at which Wittgenstein's system makes contact with epistemology, and hence (arguably) with realism. Elucidations succeed in picking out a and not b because a and b have, in the actual world, different (external) properties. This point is therefore coeval with the adoption of a stance regarding the relation of language and world that is, after a fashion, realistic. If 'a' and 'b' are names of the same form, my use of 'a' means a rather than b only because I am acquainted with a and hence can distinguish it from b. In explaining this point Wittgenstein uses the German *kennen*: perhaps he intends this as a translation of Russell's notion of acquaintance. Elucidations, Wittgenstein is therefore saying, [Wittgenstein, 1961, 3.263] can only be understood by acquaintance with

the objects in question.

3 PROPOSITIONS

3.1 Sense

We have now seen how Wittgenstein conceived of the world, as consisting of all the occurrent states of affairs. Let us turn now to his conception of how we represent this world. The fundamental idea is that this is not a matter merely of describing the world — listing, that is to say, all the states of affairs that occur. If we are to represent the world, it is essential that we encompass not only how the world is but also how it could be but isn't. What is expressive, in other words, is not simply saying that the world is a certain way, but saying that it is this way in contrast to some other way it could have been. The purpose of this chapter will be to explain this idea of representation in more detail and to sketch how Wittgenstein conceived of the structure of languages fashioned to achieve it.

For each way that objects *could* be arranged, they may or may not actually be arranged that way: each state of affairs, that is to say, may or may not occur. The world, Wittgenstein says, is just the totality of those states of affairs that *do* occur [Wittgenstein, 1961, 2.04]. So if there are N states of affairs, there are 2^N possible worlds, one for each possibility of existence and non-existence of states of affairs [Wittgenstein, 1961, 4.27]. A propositional *sense* consists in agreement or disagreement with possibilities of existence and non-existence of states of affairs: [Wittgenstein, 1961, 4.2] in other words, it divides all possible worlds into two classes and agrees with all those in one class, disagrees with all those in the other. For N states of affairs, there are 2^{2^N} such divisions [Wittgenstein, 1961, 4.42]. Notice, though, that among these divisions there are two that are somewhat trivial: that of agreement with all possible worlds, which Wittgenstein calls *tautological*; and that of disagreement with all, which he calls *contradictory*. However, their very triviality — they are not really *divisions* at all since each leaves one of the two classes of worlds empty — leads Wittgenstein to say that they are not, strictly speaking, senses: in recognition of this we shall call them *empty* senses. If there are N states of affairs, this leaves $2^{2^N} - 2$ genuine senses.

This sort of thing looks so familiar to the modern eye that it is easy to miss the significance of Wittgenstein's definition of sense. The key point is that the range of possible senses is determined once the range of elementary propositions is determined: there is no further reference to the symbolism in specifying the range. This is in contrast with Russell, whose notion of a proposition had originally been tied to the symbolism. The significance of this is that Wittgenstein's account can avoid the paradoxes by means of a *simple* theory of types: propositional functions have to stand in hierarchies, so that a function of a function of individuals is of a different type from a function of individuals; but there is no need to stratify functions any more than this, or to stratify propositions at all.

3.2 Expressing sense

At the outset I said that Wittgenstein took the purpose of propositions to be the expression of what is or could be the case. We are now in a position to make this rough schema more precise: the purpose of a proposition is the expression of a sense, i.e. the expression of agreement or disagreement with truth-possibilities of elementary propositions. Every elementary proposition is thus a proposition: it expresses agreement with just those possible worlds in which the state of affairs which it pictures actually occurs. But, as we have seen, even if there are only finitely many states of affairs, there are far more senses than the elementary propositions could express. Wittgenstein's task now is thus to provide a means for expressing these other senses.

The procedure Wittgenstein adopts is to break the task down into two parts. The first part is to devise a symbol whose purpose is to pick out some class of propositions. Wittgenstein calls this symbol a *variable* and typically denotes it ξ. He then uses the notation $(\overline{\xi})$ for the class of propositions which the variable picks out. Then he posits an operator N which can be combined with any such symbol for a class of propositions to obtain a symbol $\mathrm{N}(\overline{\xi})$ expressing the joint denial of all the propositions in the class. The sense of $\mathrm{N}(\overline{\xi})$, in other words, is such that it is true in just those worlds in which all the propositions in $(\overline{\xi})$ are false.

So far, though, Wittgenstein's description of his formal language is merely programmatic. The question how expressive it is, i.e. which of the possible senses are expressible by means of propositions in the language, depends entirely on which variables he permits for specifying classes of propositions. Wittgenstein sometimes talks as if his system will be able to express *all* possible senses, [Wittgenstein, 1922, 4.5] And indeed if there are only finitely many elementary propositions, then *every* sense is a finite truth-function of the senses of elementary propositions, and hence, as we shall see in a moment, expressible in Wittgenstein's notation using the simplest and most obvious method, direct enumeration. But in the case when there are infinitely many elementary propositions this will plainly not be possible, at least if we restrict ourselves to the use of finite symbols, since there will be uncountably many possible senses but only countably many finite symbols in the formal language. What this shows is, roughly speaking, that the expressive power of the system is determined not by the N operator but by the range of classes of propositions that the variable notation permits us to pick out.

Wittgenstein specifies three types of variable explicitly, and evidently supposes that they will be adequate for all the practical purposes that will normally arise within the language. They are

> 1. Direct enumeration. ...
> 2. Giving a function fx, whose values for all values of x are the propositions to be described.
> 3. Giving a formal law that governs the construction of the propositions. [Wittgenstein, 1961, 5.501]

In the next few sections we shall consider these types in turn, and then address the question whether there are any other types of variable that Wittgenstein needs.

3.3 Direct enumeration

One way of picking out a class of propositions, at least if the class is finite, is simply to list them. If $p_1, p_2, \ldots, p_n$ are propositions, Wittgenstein uses the notation $(p_1, p_2, \ldots, p_n)$ for the class whose members are just these propositions. When we combine this symbol with the N operator, we obtain the expression $\mathrm{N}(p_1, p_2, \ldots, p_n)$: it represents the truth-function of $p_1, p_2, \ldots, p_n$ which is true just in case all of $p_1, p_2, \ldots, p_n$ are false, so it would be expressed in Russell's notation as $\sim p_1 \cdot \sim p_2 \cdot \ldots \cdot \sim p_n$.

Note in particular that when the class has two members this reduces to

$$\mathrm{N}(p, q) = \sim p \cdot \sim q = p|q,$$

where $|$ is the truth-function sometimes referred to as the 'Sheffer stroke'. The reason this is important is that, as Sheffer had proved in 1913, the stroke is adequate to express *every* finite truth-function. (In fact, Sheffer proved that there are two binary truth functions each of which is adequate: the other one — also sometimes called the 'Sheffer stroke' — is $\sim p \vee \sim q$.) Since the stroke is expressible using the N operator together with the notation for direct enumeration, it follows straightaway that Wittgenstein's notation is also adequate to express every finite truth-function.

3.4 Functions

As we have seen, in the finite case the method of direct enumeration enables us to express every sense, since it gives us all the expressive power of the propositional calculus. But it is a familiar fact that in the infinite case the propositional calculus is not only inadequate to express every possible sense but, more importantly, inadequate even to express every sense that we have a use for. What we need if the symbolism is to be adequate for practical purposes is to effect the same extension of the expressive power of the system that Frege and Russell achieved by adding a notation for quantification. The way Wittgenstein achieved this was by allowing a propositional function to count as a variable picking out the class of propositions that is its range.

The idea is in essence very simple. Take some sign expressing a proposition and single out part of the sign. Now keep this part fixed and let the rest of the sign vary. All the propositions that can be obtained by varying the sign in this way form a class. The variable used to pick out this class is called a propositional function, but Wittgenstein invariably refers to it simply as a *function*.

The part of the proposition that is taken as fixed in forming a propositional function plainly cannot be arbitrary. It would not make much sense, for example, to treat 'quickly into' as fixed in the proposition 'John walked quickly into the

room'. The part that is fixed has to be a unit that contributes significantly to the way in which the sign as a whole expresses a proposition. The most familiar example is the case of a propositional sign in which a name 'a' occurs. From such a sign we can obtain a function '$f(x)$' whose instances are all the propositions obtained by substituting various names in the position where 'a' occurred originally.

Wittgenstein does not introduce an explicit notation for the class of all such propositions: here I shall write it $\widetilde{x}f(x)$, cf. [Geach, 1981]. With this notation we can obtain, for instance,

$$\begin{array}{rcl} \mathrm{N}(\widetilde{x}fx) & = & \sim(\exists x)fx \\ \mathrm{N}(\mathrm{N}(\widetilde{x}fx) & = & (\exists x)fx \\ \mathrm{N}(\widetilde{x}\mathrm{N}(fx)) & = & (x)fx. \end{array}$$

The notation takes care of multiple generality. For instance,

$$\begin{array}{rcl} (x)(\exists y)xRy & = & \sim(\exists x)\sim(\exists y)xRy \\ & = & \mathrm{N}(\widetilde{x}\mathrm{N}(\widetilde{y}xRy)). \end{array}$$

The fact that Wittgenstein does not explicitly introduce any notation for the range of a propositional function has perhaps contributed towards a belief among some commentators that his system cannot express multiple generality, but as we have seen, this is wrong. (If it were indeed the case that the system did not suffice to express multiple generality, it would be at the very least surprising that neither Russell nor Ramsey noticed it.)

The sort of propositional function we have been considering so far is not the only one that is possible, though. Wittgenstein's system also permits higher order quantification. In a proposition in which a name 'a' occurs, instead of keeping the rest of the proposition fixed as we did before, we can keep a fixed and let the rest vary. That way we obtain a variable $\phi(a)$ specifying a class of propositions in the expression of which the name 'a' occurs.

At this point things becomes less clear, however. Suppose, for instance, that $f(a)$ is elementary. When we kept the remainder of the proposition fixed in order to form the function $f(x)$, it was clear what the range of x was — namely all those names which can meaningfully be substituted for x. But now that we are keeping a fixed and letting the remainder vary to form $\phi(a)$, it is less clear what the range of variation is. One possibility is that $\phi(a)$ should range over all the elementary propositions of the same form as $f(a)$. Another, at first sight anyway, is that $\phi(a)$ range over *all* the propositions in which 'a' occurs. The difficulty is that this includes propositions in the expression of which this very variable ϕ occurs. Can Wittgenstein allow this without falling into paradox? This is a complex issue and we shall postpone discussion of it to a later section.

3.5 Formal series

Consider the following series of propositions:

$$\begin{array}{c} aRb \\ (\exists x) : aRx \cdot xRb \\ (\exists x, y) : aRx \cdot txRy \cdot yRb \\ \vdots \end{array} \qquad (*)$$

For various purposes it is useful to have a variable which ranges over all the propositions in such a series. For instance, if ξ is such a variable then $\mathrm{N}(\mathrm{N}(\overline{\xi}))$ will express the weak ancestral of R, i.e. the relation between a and b which holds if at least one of the propositions in the formal series obtains.

Now the conventional method of achieving this is to write aR^nb as an abbreviation for the nth term of the formal series, i.e. for

$$(\exists x_1, \ldots, x_{n-1}) : aRx_1 \cdot x_1Rx_2 \cdot \ldots \cdot x_{n-1}Rb.$$

If we do this, then in the notation introduced in the previous section $\widetilde{y}\, aR^yb$ is the variable we require. But is the transition from the proposition aR^nb to the variable $\widetilde{y}\, aR^yb$ permissible? For it to be so, the 'n' in aR^nb has to contribute significantly to the way in which this symbol expresses its sense. But in Wittgenstein's view it does not do so: indeed natural numbers *never* occur in propositions as arguments of propositional functions.

Wittgenstein does not give an explicit argument for this in the text, though, or even state it distinctly, so enunciating his reason for holding it involves some reconstruction. If it were not so — if, in other words, we could form a function $f(n)$ with a natural number variable — then 0, 1, 2, etc. would all have to be expressions of the same type. It would not follow from this that the numbers would themselves have to be Tractarian objects: one could envisage, for instance, an account on which numerals were complex expressions analysable into something simpler. But whatever occurred at the ultimate level in this analysis *would* be Tractarian objects. These, which we could reasonably call *mathematical* objects, would constitute in a sense the subject matter of mathematics.

But Wittgenstein did not think that mathematics *can* have a subject matter in this sense any more than logic can.

> If there were mathematical objects — logical constants — the proposition 'I am eating five plums' would be a proposition of mathematics. And it is not even a proposition of applied mathematics. [Wittgenstein, 1979, 14 Feb. 1915]

We can see that it is 'not even a proposition of applied mathematics' simply by rewriting it so that 'five' does not occur in it. So nothing characterizes propositions in which number words occur. It follows that mathematics does not have a subject matter in the sense required.

So the method considered above of forming a variable to pick out the propositions occurring in a formal series is not available. Wittgenstein was therefore forced to introduce a new kind of variable to do this job. His notation depends on the observation that a formal series is determined by two things, an initial class of propositions already fixed by a variable ξ, and an operation O, that is to say a method of obtaining new propositions from old. Wittgenstein then combines these materials to form the notation $[\overline{\xi}, x, Ox]$ to express a variable ranging over all the propositions in the class $(\overline{\xi})$ together with those obtainable from them by repeated application of the operation O. Thus, for example, the formal series is expressed by the variable

$$[aRb, a\chi b, (\exists x) : aRx \cdot x\chi b].$$

As Wittgenstein himself emphasizes, operations and functions must not be confused with one another. The key difference is that an operation, unlike a function, is repeatable. This repeatability is evidently central, since it is what enables the variable $[\overline{\xi}, x, Ox]$ to generate successively more propositions.

Now the reason a function is not repeatable is, roughly speaking, that the argument of the function occurs as a symbol in the symbol for its value. Thus if the function is fx, the symbol 'fa' for the value contains the symbol 'a' for the argument. Nothing corresponding occurs in the case of an operation, as is especially obvious in the example we considered above of the operation which generate successive terms of the formal series: it is plain, for instance, that the symbol aRb does not occur in the symbol $(\exists x) : aRx \cdot xRb$.

This last point is of considerable significance because there are other examples, for which it is much less obvious. It would be easy, for example, to be fooled into thinking that 'p' occurs in the symbol '$\sim p$', so that $\sim p$ is a function of p. If this were so, negation would not be repeatable. We shall return later to the significance of this observation for Wittgenstein's system.

3.6 The general form of proposition

At the extreme it is possible to regard a single proposition p as a variable, even if a rather trivial one. [Wittgenstein, 1961, 3.313] Indeed this can be done using each of the three methods considered so far: we can directly enumerate the class (p) whose only member is p; we can regard p as a propositional function in which the whole proposition is held constant and no part of it varies, so as to obtain the class $\widetilde{x}p$; or we can take the variable $[p, x, x]$, which expresses the formal series obtained from p by repeatedly applying the trivial operation that takes any proposition to itself.

It is natural to ask, then, whether at the other extreme there is a variable ranging over *all* propositions. If there were only finitely many states of affairs, of course, we could list expressions for all possible senses by direct enumeration. This is worth noting, because if there is a variable ranging over all propositions in the finite case, there cannot be a general theoretical ban on such a variable.

Nevertheless it does nothing to help us construct one in general. In the infinite case direct enumeration obviously will not work, so let us consider instead the propositional function method.

To do it by this method we would need to be able to form a propositional function that could take any proposition at all as its value. Notice that for this it is not sufficient merely to be able to let *every* part of a proposition vary and keep no part fixed. If, for example, we let every part of the proposition $Fa \vee aRb$ vary, we could obtain a variable, which we might write $Xx \vee yYz$, ranging over propositions expressible as disjunctions of propositions of the same form as Fa and aRb, but it is not at all clear how we could obtain *all* propositions by this means.

What, then, of Wittgenstein's third method? This gives us the expression $[\overline{p}, \xi, \mathrm{N}(\overline{\xi})]$ for the general form of proposition, where $\overline{p}$ is a variable ranging over the elementary propositions. So we have now reduced the problem from that of finding a variable that ranges over *all* propositions to that of finding one that ranges over the elementary ones. But is this problem any easier? We still have the same difficulty over the variability of form. If we take a particular elementary proposition $a_1a_2 \ldots a_n$ and turn every name in it into a variable, we get $x_1x_2 \ldots x_n$, which is a prototype for all elementary propositions of this type. But there may very well be other elementary propositions of some other type, such as $b_1b_2 \ldots b_m$. The prototype for this would be $y_1y_2 \ldots y_m$, which is not the same. So how do we get one variable to range over both types?

We could just enumerate the prototypes, which would at any rate work if there are only finitely many types, but this appears rather too accidental. We seem forced to say that a variable ranging over all elementary propositions could not conform to any of the three kinds we have yet considered, and hence that if there is to be such a variable, 'elementary proposition' must be a distinct formal concept.

4 THE LIMITS OF LANGUAGE

4.1 Sign and symbol

In the course of our discussions we have had cause to mention various linguistic items. But there are several things we might have been referring to in these discussions. First, of course, there is a particular string of ink marks on a page, i.e. the token: and it is a logical commonplace to distinguish this from the type of which all similar strings of marks are tokens. For Wittgenstein a *sign* is the type, not the token: hence the wonderfully compressed remark in the *Tractatus* that ' "A" is the same sign as "A".'[Wittgenstein, 1961, 3.203] Notice, though, that in order to identify the type, we need a criterion to settle which inscriptions count as similar, and such a criterion is applicable only to facts, not complexes. Thus a complex sign '*abc*' is for Wittgenstein not the complex entity made up out of the individual signs '*a*', '*b*' and '*c*'; it is the *fact* that these signs are arranged in this way. Signs, in other words, are to be individuated not merely according to the

objects they are made up of, but according to the relationships that hold between those objects. A sign is thus an inscription parsed in such a way as to exhibit its grammatical structure. But the *grammatical* structure here consists only in what can be expressed purely in terms of rules for signs without bringing in any considerations of meaning. As soon as we read the sign as having a meaning — as having one meaning rather than another — we have turned it into what Wittgenstein calls a *symbol*; we have placed it in a relationship not merely to other linguistic items but to the world.

The distinction between sign and symbol is thus nicely placed to help explain the confusions that arise in natural languages when the same sign should properly be read in different occurrences as a different symbol. Such confusions would be avoided if we could adopt a 'logically perfect' language, since in such a language each sign has a unique reading. But the distinction between sign and symbol is of much greater significance than this, for it is in the transition from one to the other that Wittgenstein locates the self.

Establishing the plausibility of this claim will take us some time. Notice first that if it were not for the role that it plays in his account of the self, Wittgenstein would have had no need to draw the distinction between sign and symbol in the case of logically perfect languages. The evidence of the *Prototractatus* bears this out, for we can identify quite precisely there the point at which Wittgenstein first introduces this terminological distinction. It occurs on page 53 of the manuscript in Wittgenstein's explanation of his remark that if what appears to be a propositional sign lacks sense, that can only be because we have failed to give a meaning to one of its constituents. He illustrates this with a case where a sign symbolizes in a different way in two occurrences and introduces a distinction between signs (*Zeichen*) and their visible part (*ihren sichtbaren Teil*) to explain this [Wittgenstein, 1971, 5.30641]. Then, in the very next remark on the page, he replaces this with the distinction between *Symbol* and *Zeichen* that is now familiar. He first of all introduces it for the special case of simple signs: 'Das einfache Zeichen ist der wahrnehmbare Teil des Namens.' Then he at once generalizes to signs in general: 'Das Zeichen ist der sinnlich wahrnehmbare Teil des Symbols.' [Wittgenstein, 1971, 3.2013]. And once he has the distinction for signs in general he can apply it in particular to distinguish the propositional sign (*Satzzeichen*) from the proposition (*Satz*).

Wittgenstein's use of these words before page 53 does not prefigure the terminological distinction drawn there. Before that point he uses *Zeichen* indifferently for what he now distinguishes as *Zeichen* and *Symbol*, and *Satzzeichen* for what he now distinguishes as *Satzzeichen* and *Satz*. We know that Wittgenstein did not draw the terminological distinction before page 53 because some occurrences of *Zeichen* before that point become *Symbol* in the final version of the book and others do not; in the same way most uses of *Satzzeichen* change to *Satz* but crucially one — the remark that a propositional sign is a fact[1] — remains unchanged.[2]

[1] [Wittgenstein, 1971, 3.16, 3.161], [Wittgenstein, 1922, 3.14, 3.143].

[2] So Sundholm's claim in [Sundholm, 1992] that in the *Prototractatus* the word *Satzzeichen*

Wittgenstein did not make these changes at once, though: he did not immediately work through the first 53 pages of his manuscript looking for occurrences of *Zeichen* to change to *Symbol*. With one exception he left these revisions to a later draft of the book. The exception is pages 24 and 25, where he altered a series of consecutive remarks[Wittgenstein, 1971, 3.20121, 3.20122, 3.22, 3.23 and 3.24] to accord with his new terminology. The reason he changed only these remarks is easy to see: when he wrote on page 53 the proposition introducing the *Symbol* terminology he had to look back to page 24 to determine what its number ought to be[Wittgenstein, 1971, 3.2013] and took the opportunity of revising the adjacent propositions straightaway.

Until he came to the discussion of ambiguity, then, Wittgenstein did not in the *Prototractatus* draw any terminological distinction between sign and symbol, using *Zeichen* indifferently for both notions. It is natural to wonder, therefore, whether he was even aware of the difference before that. This is evidently important in the present context: if he was not aware of the difference until half way through writing the *Prototractatus*, that would be evidence against locating the self in the gap between the two notions. In fact, though, Wittgenstein did have the distinction much earlier: the remark that in the final text is rendered, 'Um das Symbol am Zeichen zu erkennen, muss man auf den sinnvollen Gebrauch achten' [Wittgenstein, 1922, 3.326] began life in the first wartime notebook as 'Um das Zeichen im Zeichen zu erkennen, muss man auf den Gebrauch achten',[Wittgenstein, 1979, 23 Oct. 1914] which is quite inexplicable unless Wittgenstein already attached two distinct senses to the word '*Zeichen*'.

If, as McGuinness [1989] has conjectured, Wittgenstein wrote the first 70 pages of the *Prototractatus* after the second wartime notebook but before the third, we should not expect the use of the word *Symbol* defined on page 53 to occur in the first two notebooks. Nor do we. Indeed it does not occur in the third notebook either, but this does not create an argument for a later dating of the *Prototractatus* than McGuinness recommends, since none of the occurrences of *Zeichen* in the third notebook are obviously in error once the distinction between *Zeichen* and *Symbol* is in place.[3] A potential source of confusion here is that in the two pre-war texts dictated by Wittgenstein (to Russell and to Moore), both of which are in English, the word used is 'symbol'. This indicates no more than that at that time, without any terminological distinction on the horizon, Wittgenstein was happy to license 'symbol' as an acceptable English translation of '*Zeichen*' (which it is).

When he started writing the *Prototractatus*, then, Wittgenstein did not have the terminology that later enabled him to distinguish between a propositional sign (a species of sign) and a proposition (a species of symbol). But he did already have another terminological distinction that could play a similar role, namely that

is used systematically for what in the *Tractatus* is called *Satz* is not accurate: it is in fact used indifferently for both notions.

[3]An argument could be mounted that Wittgenstein's remark of 22nd November 1916 — 'The concept of the operation is quite generally that according to which signs can be constructed according to a rule' — would make more sense with 'symbols' rather than 'signs', but it is probably not a strong enough argument to overturn McGuinness' speculations.

between what he *then* called a propositional sign and a proposition. '*Satzzeichen*' in the terminology prior to page 53, let us recall, is ambiguous between what are later called a propositional sign and a proposition: it is a sensibly perceptible fact whose structure fits it to the task of expressing a sense. A *Satz*, as the word is used before page 53, is such a fact together with the manner of its depiction of what it expresses, [Wittgenstein, 1971, 3.2] i.e. together with the projection that maps its sense (if it has one) onto it.

It is this last notion of a proposition with its mode of depiction somehow attached to it that drops out in the later terminology: what Wittgenstein's discussion of ambiguity seems to have shown him is that he simply did not need it. What is it that we do when we recognize the symbol in a sign? If the sign is a propositional sign, we recognize the projection that maps its sense onto it (its mode of depiction). The mode of depiction is thus implicit in the symbol. There is, once Wittgenstein has made his new terminological distinction, no need for the old one, and what was then called a *Satz* (a proposition together with its mode of depiction) is henceforth simply dropped from the account.

4.2 Intensional contexts

Wittgenstein had to face the problem that our language apparently contains intensional contexts like 'A believes that p'. He analysed this as being of the form '"p" says that p'. [Wittgenstein, 1961, 5.542] But what does that mean? Before we answer this question, let us put it in context by examining the views held by Russell on the same issue up to and during his interaction with Wittgenstein.

Until 1907 Russell thought that propositions were complexes actually contain the things of which they spoke, so that the snow-covered peak itself is a constituent of the proposition that Mont Blanc is the tallest mountain in Europe. The difficulty with this is that it does not account for false propositions. The proposition that Charles I died in bed cannot be identified with the fact of his doing so, because there is no such fact. And we cannot simply suppose that there are negative facts as well as positive ones, since 'this leaves the difference between truth and falsehood quite inexplicable'. [Russell, 1910, p152] At this point Russell applied an insight he had learnt from 'On Denoting', that the sentences of a language might be superficially misleading. It might be that expressions occurring in English sentences, for example, vanished on analysis, from which one could infer that they do not correspond to anything in the world. Thus, for example, 'The King of France' vanishes when you analyse 'The King of France is bald' in the familiar Russellian manner. Applied to the present question, that insight amounts to this: it might be that the term apparently denoting the proposition vanishes from the correct analysis of statements attributing propositional attitudes.

So Russell tried to analyse a propositional attitude ascription as a relation holding between the bearer of the attitude and the components of the (apparent) proposition. Thus, for instance, one might as a first attempt try to analyse 'A judges that b loves c' as $J(A, b, \text{loving}, c)$, where J is the judging relation, b and

c are the putative lovers, and *loving* is the universal, i.e. the abstract relation of loving.

But this cannot be quite right. A more refined account states the analysis as $J(A, b, \text{loving}, c, x\chi y)$, where the last term denotes the general form of binary relations and indicates that A judges b to be the lover and c the beloved.

There are two difficulties with this conception though. First, there is a difficulty connected with the existence of logically complex judgments. Russell's view was that 'A judges that p or q' would have an analysis of the form $J(A, p, \vee, q, F)$, where F denotes the form of the proposition and we are ignoring for the moment the fact that in a fuller analysis p and q would themselves be broken down into their constituents. Russell also thought that a full analysis would display how the proposition reaches right up to reality, that is, that it would only contain genuinely denoting expressions. It is a consequence of this that the logical constants really do denote objects in the world that can be known by acquaintance. (In fact for Russell, though not for Wittgenstein, it was more complicated than this because in his theory of types the logical constants display typical ambiguity: one would have to say that there is a distinct '$\vee$' for each type.)

The problem, though, is that the logical constants are interdefinable [Wittgenstein, 1961, 5.4–5.42]. For example $p \vee q$ can be defined as $\sim (\sim p. \sim q)$. Alternatively you could define $p.q$ as $\sim (\sim p \vee \sim q)$. This shows that any particular logical constant can be made to vanish through suitable analysis. There are no grounds for saying that any one of them is more fundamental than any other — no grounds, therefore, for determining which of them is logically simple and known to us by acquaintance, which logically complex and known to us only by description.

Perhaps Russell could have met this objection. But a more serious objection arises from the fact that by including the verb of the proposition judged as a term in the judging relation, Russell had to turn it into a noun, therby removing from it its ability to ensure the grammaticality of what is judged. Or, better, if J can have arguments, it cannot have argument *places*. In particular, the analysis of 'A judges that the table is to the right of the book' is such that the position occupied therein by the relation term 'right' could equally sensibly be filled by the predicate 'penholder'. So it would make equally good sense to say 'A judges that the table penholders the book' — be equally legitimate, that is, to ascribe to A the judgment of a piece of nonsense to A. Wittgenstein inferred that the theory is wrong, because 'The correct explanation of the form of the proposition, "A makes the judgment p", must show that it is impossible for a judgment to be a piece of nonsense' and 'Russell's theory does not satisfy this requirement.' [Wittgenstein, 1961, 5.5422].

Wittgenstein's approach, by contrast, is a consequence of his picture theory of meaning [Wittgenstein, 1961, 2.1 ff.]. The idea behind this was, as we have seen, that a proposition has the capacity to express a fact in virtue of its standing in a pictorial relation to the world. In a picture, each element of the picture is correlated with a thing, and that these elements are arranged in a certain way says that the corresponding things are arranged in the same way. It is the same with

a proposition: [Wittgenstein, 1961, 3.11, 4.032] the elements of the proposition (the names) are correlated with things (the simple objects); that the names are arranged in a certain way says that the correlated things are arranged in that way [Wittgenstein, 1961, 3.14]. The only difference between a picture and a proposition is that the pictorial form of the latter is more general.

When Wittgenstein says that 'A believes that p' is of the same form as ' "p" says that p', he is analysing it as: 'p' is a picture of p. A picture (or a proposition) is itself a fact [Wittgenstein, 1961, 2.141]. Of course a picture can take many forms without losing what is essential to its capacity to depict [Wittgenstein, 1961, 4.011–4.016] and among these forms are the mental states corresponding to propositional attitudes. In particular, Wittgenstein regards A's belief that p as a picture, in a mental medium, of p; it is, that is to say, an arrangement of mental objects, to which the expression ' "p" ' refers.

For Wittgenstein, ' "p" says that p' is therefore one of those things that can be shown but cannot be said, because it cannot express something true and significant. It is trying to say, in language, how a bit of language is connected with the world. That would be like trying to display, on a map, the method of projection whereby a bit of the map depicted formations on the ground. Any attempt to depict it on the map can never be informative, because to grasp it one would already have to grasp the method of projection [Wittgenstein, 1961, 4.022]. There must be some formal community between the system of signs and the things that are signified. (For instance, a map symbolises a part of the surface of the Earth because they share two-dimensional spatial form.) For it is essential to language that it can express all the possibilities, and the form of a language determines what possibilities it can express, whereas the form of the world determines the possibilities that there are. So world and language must share form [Wittgenstein, 1961, 2.18-2.203]. That is why one cannot represent a geometrically impossible configuration on a map [Wittgenstein, 1961, 3.032-3.0321]. But, as in the case of the map, this form cannot itself be conveyed in the language, for an understanding of the language presupposes a grasp of the form of language (which is also the form of the world); and this can only be shown [Wittgenstein, 1961, 4.12–4.121]. It is a (surprising) consequence of all this that you cannot really *say* that A believes that p.

4.3 The self, value and the will

This analysis, unlike Russell's, makes no mention of the apparent judging subject A: analysis has shown 'A' not to be a genuinely denoting term. 'This shows ... that there is no such thing as the soul — the subject, etc. — as it is conceived in the superficial psychology of the present day.' [Wittgenstein, 1961, 5.5421]. Wittgenstein's analysis [Wittgenstein, 1961, 5.542] of 'A believes that p' shows that there is no simple self that is the owner of some particular set of psychological states among others, but distinct from any of them. But in the next remark Wittgenstein also dismisses the idea of a composite soul. This amounts to rejecting the idea that we could identify the self with any collection of mental states, and hence with

some complex of states of affairs that might or might not obtain.

Wittgenstein explains his rejection of that idea by means of an analogy between the self and the seeing eye [Wittgenstein, 1961, 5.6ff]. Just as the eye is not part of visual experience, neither can the self take part in expressible states of affairs. The eye that looks on the world is not a part of the visual world. Nothing in the visual field tells us that there is an eye that "has" the visual field. It is true that when you look in a mirror you see an eye. That is not the grammatical or seeing eye, however, but merely a complex physical organ. Similarly you can attempt to describe a part of the world of states of affairs, but you can never put the I that does the describing into that description. We should think of the self as a limit of the world, not a part of it.

Here also [Wittgenstein, 1922, 5.64] Wittgenstein remarks that solipsism coincides with 'pure realism'. If we think of the self as outside the world, representing it, then the sceptical and idealistic possibilities arise. Why, for example, should we not suppose that there is no world at all, just the self and its representations? The realist, on the other hand, subverts this picture and instead conceives of the self, not indeed as part of the world, but as the limit of the world [Wittgenstein, 1961, 5.632]. Here 'limit' means the possibilities that are open to it, i.e. for Wittgenstein the possible configurations of objects. But this picture is inevitably solipsistic too. If the self is just the set of possibilities open to the world, then there can only be one self, for there is only one set of possibilities open to the world. 'The world is my world: this is manifest in the fact that the limits of language (of that language which alone I understand) mean the limits of my world.' [Wittgenstein, 1961, 5.62].

There is a connection between Wittgenstein's treatment of the self and his treatment of aesthetic and ethical value (the two being identified at 6.421). The difficulty is to see how value finds a place in the system of the *Tractatus*. Propositions can express only matters of fact; and whatever is expressed by saying that a particular painting is beautiful, for instance it cannot be a mere matter of fact [Wittgenstein, 1961, 6.41]. For if it were, we would thereby recognize the possibility that it is not beautiful, and then it would only be chance that the actual world is one in which the object is beautiful.

This in turn is related to the question of agency. The metaphysical self has a will, as Wittgenstein says, [Wittgenstein, 1922, 5.631] and yet is not part of the world:[Wittgenstein, 1922, 5.632] how then can my willing bring about anything in the world?

Wittgenstein distinguishes two kinds of will: the psychological phenomenon of volition, and the will of the transcendental self, the latter being the locus of value [Wittgenstein, 1961, 6.423]. He implies that in neither sense can the will bring about anything else in the world [Wittgenstein, 1961, 6.37, 6.373–4]. The world is a world of facts. Space and time are constructed from simple objects just as everything else is. In the empirical sense our ascriptions of change to things will just amount to certain situations being the case. But he thinks that good willing (in the transcendental sense) changes the world in another way, that is, by altering

the limits of the world. 'The effect must be that it becomes an altogether different world. It must, so to speak, wax and wane as a whole.' [Wittgenstein, 1922, 6.43]

One might try explaining this remark as follows: good willing enriches the world by adding new objects to the world as conceived by the subject; bad willing impoverishes it by removing some objects. That is, good willing refines the subject's language: its stock of simple names is enlarged. In view of Wittgenstein's holism with regard to fully analysed languages, this amounts to the subject's acceding to a wholly new language and correspondingly a wholly different world. 'The world of the happy man is a different one from that of the unhappy man.' The world of the former is open to more possibilities than that of the latter.

But this interpretation is in tension with Wittgenstein's observation that 'if good or bad willing changes the world, it can alter only the limits of the world, not the facts; not what can be expressed by means of language.' This remark suggests an alternative interpretation according to which good willing changes not what there is in the world but what its significance is. The world of the happy (i.e. the good-willing) man is of greater importance than that of the unhappy man: it glows more brightly, as it were. The fundamental activity of the self is the infusion of dead signs with symbolic meaning, and thus the expression of the world. For the happy man this activity is necessarily momentous; the unhappy man, by contrast, could not care less whether his words mean anything. What is striking now is how rich Wittgenstein's conception of the limits of the world has become. On this second interpretation the world of the happy man is expressed in the same language as that of the unhappy, and hence allows for the same range of possibilities; yet it is correct to describe it as 'an altogether different world' since its significance is different.

5 THE THEORY OF TYPES

5.1 Higher types

In 4.1 we went on a rather lengthy excursion through the origins of Wittgenstein's distinction between signs and symbols, and in particular between propositional signs and propositions. What that excursion demonstrated was that there is nothing in those origins to dissuade us from giving the sign-symbol distinction a central role in Wittgenstein's system. What we need to do now is to see how the distinction enters into Wittgenstein's attempted resolution of the logical paradoxes.

A mention in David Pinsent's diary on 25th October 1912 [von Wright, 1990, p37] of Wittgenstein's solution to 'a problem which has puzzled Russell and Frege for some years' may well be a reference to Russell's paradox, but even if it is, there is no need to suppose that the solution we find in the *Tractatus* is exactly the one Wittgenstein had in 1912. (His letters to Russell from that period hint at experiments with logical theories unrecognizable to a reader of the *Tractatus*.) On the other hand, the solution in the *Tractatus* is not in its bare outlines particularly original.

> The reason why a function cannot be its own argument is that the sign for a function already contains the prototype of its argument, and it cannot contain itself.
>
> For let us suppose that the function $F(fx)$ could be its own argument: in that case there would be a proposition '$F(F(fx))$', in which the outer function F and the inner function F must have different meanings, since the inner one has the form $\phi(fx)$ and the outer one has the form $\psi(\phi(fx))$. Only the letter 'F' is common to the two functions, but the letter by itself signifies nothing. ... That disposes of Russell's paradox. [Wittgenstein, 1961, 3.333]

Suppose that we apply this explanation in the particular case when $F(f\widehat{x})$ says that f does not hold of itself. Russell's paradox results from asking whether or not F is predicable of itself, whether $F(F(f\widehat{x}))$ in other words. But in asking this question we use F in a different sense from the old one: one sign occurs twice but is a different symbol each time. On this analysis Russell's paradox is exposed as no more than a bad pun.

The doctrine Wittgenstein is appealing to here is that how a sign for a function symbolizes in some way presupposes the structure of the argument place of the function. Wittgenstein draws the by now conventional conclusion that 'functions stand in hierarchies', but of course this is not in itself enough to render the theory secure. More needs to be said if his system is not to fall prey to what are nowadays called the semantic paradoxes. Consider, for example, Grelling's 'heterological' paradox, which is one of the best-known: an adjective is said to be heterological if it does not apply to itself; the antinomy results from asking whether the adjective 'heterological' is itself heterological. The paradox is easily formulated in the simple theory of types. Let us write '$\mathrm{Het}\, x$' for

$$(\exists\phi)(x \text{ means } \phi\widehat{z}.\sim\phi x).$$

Now

$$\text{'Het' means } \mathrm{Het}\,\widehat{z},$$

and so

$$\begin{aligned}\mathrm{Het}(\text{'Het'}) &\equiv (\exists\phi)(\text{'Het' means } \phi\widehat{z}\dot{\sim}\phi(\text{'Het'}))\\ &\equiv\sim \mathrm{Het}(\text{'Het'}),\end{aligned}$$

since 'Het' cannot mean two different things. What we need, therefore, is a way of avoiding paradoxes of this sort.

5.2 Russell's reasons for ramification

Let us consider Russell's way first. Propositions (and hence propositional functions) are stratified into orders. No variable is permitted to range over more than one order of propositions (or propositional functions). This solves the heterological

paradox by making it impossible to define the function $\mathrm{Het}\,\widehat{x}$ at all. The nearest we can get to it is to define a whole hierarchy of functions $\mathrm{Het}_1\widehat{x}, \mathrm{Het}_2\widehat{x}$, etc. by writing '$\mathrm{Het}_n x$' to mean

$$(\exists\phi_{n-1})(x \text{ means } \phi_{n-1}\widehat{z}\dot{\sim}\phi_{n-1}x).$$

We can then deduce that

$$\begin{aligned}\mathrm{Het}_n(\text{`}\mathrm{Het}_{n-1}\text{'}) &\equiv (\exists\phi)(\text{`}\mathrm{Het}_{n-1}\text{' means } \phi\widehat{z}\cdot \sim\phi(\text{`}\mathrm{Het}_{n-1}\text{'}))\\ &\equiv \ \sim \mathrm{Het}_{n-1}(\text{`}\mathrm{Het}_{n-1}\text{'}),\end{aligned}$$

but this is not contradictory. The derivation of the contradictory

$$\mathrm{Het}_n(\text{`}\mathrm{Het}_n\text{'}) \equiv \ \sim \mathrm{Het}_n(\text{`}\mathrm{Het}_n\text{'})$$

is blocked because the propositional function $\mathrm{Het}_n\widehat{x}$ is of order n and hence not among the possible values of the variable $\phi_{n-1}\widehat{x}$.

Now in Russell's conception of logic the repository in which unexplained difficulties were to be located had from the time of 'On denoting' been the variable. In that paper he famously solved the difficulties of denoting phrases such as definite descriptions by treating them as contextually defined expressions that disappear on analysis. The one exception to this was the variable, which remained as a denoting concept that resisted analysis and had therefore to be logically simple. In 1906 Russell briefly thought that he could solve all the paradoxes by means of a substitutional theory that treated propositional functions too as contextually defined expressions that disappear on analysis in much the same way as definite descriptions. The virtue of this theory was that it allowed him to retain an *unrestricted* conception of the variable as ranging over all entities. When he realized that the theory did not avoid all the paradoxes, he was forced to adopt the hierarchy of orders of proposition just described and to abandon finally a single unrestricted variable. It then became the central challenge for his account to explain what the relationship was between the various variables his theory now required. The difficulty was that the various variables had all to be logically simple, so the difference between them could not be a feature of their internal structure and had therefore to consist in a logical relation between them. But such a logical relation would have to involve an unrestricted variable, which is just what the theory said was impossible. Now Russell's conception of logic was thoroughly impersonal: logic was not part of our apparatus for understanding the world, but part of the world itself. In particular, Russell saw the variable not as part of our notation for describing the world, but as a logical part of the world itself. So the difficulty that the hierarchy of variables presented to him was not the fragmentation of our relationship with the world, but the fragmentation of logic itself.

5.3 Wittgensteinian propositions

For Wittgenstein, on the other hand, the difficulty was one of configuring the self. We represent the world by means of language. Our representations are

constrained by the structure of the general form of proposition. If propositions fall into a hierarchy of orders, there is not one general form of proposition but infinitely many. It then becomes problematic how we can conceive of a single finite intelligence as possessing such a structure.

Through most of the wartime notebooks Wittgenstein seems to have accepted that propositions fall into orders, but at the end of 1916 he began to address directly the difficulty this poses:

> The fact that it is possible to erect the general form of proposition means nothing but: every possible form of proposition must be FORESEEABLE.
>
> And that means: We can never come upon a form of proposition of which we could say: it could not have been foreseen that there was such a thing as this.
>
> For that would mean that we had had a new experience, and that it took that to make this form of proposition possible.
>
> Thus it must be possible to erect the general form of proposition, because the possible forms of proposition must be a priori. Because the possible forms of proposition are a priori, the general form of proposition exists.
>
> In this connexion it does not matter at all whether the given fundamental operations, through which all propositions are supposed to arise, change the logical level of the propositions, or whether they remain on the same logical level.
>
> If a sentence were ever going to be constructable it would already be constructable. [Wittgenstein, 1979, 21 Nov. 1916]

Wittgenstein asserts here that it must be possible to erect the general form of proposition, but he does not say what that general form is. As long as he adhered to a theory which stratified propositions into orders to avoid paradox, he had to assume that this general form is inexpressible. For if there is a single operation 'N' that generates all propositions, and if $\overline{p}$ represents the class of all elementary propositions, then $[\overline{p}, \xi, \mathrm{N}(\overline{\xi})]$ represents all propositions. But in that case quantification over all propositions is possible, which is just what we wanted to avoid. If propositions are to remain stratified, we have to assume instead that there is a series of operations $\mathrm{N}_1, \mathrm{N}_2, \mathrm{N}_3, \ldots$ which generate the propositions of orders 1, 2, etc. successively from the elementary propositions.

Suppose that we agree to use one sign 'N' for all the operations $\mathrm{N}_1, \mathrm{N}_2$, etc. In that case we have an example of just the sort of ambiguity Wittgenstein's distinction between signs and symbols was designed to help explain. The sign 'N' in its various occurrences can act as infinitely many different symbols. If $\mathrm{N}_1, \mathrm{N}_2$, etc. are distinct operations with nothing in common, it follows that the way 'N' symbolizes in each of its occurrences is different. In that case we cannot grasp

all of these ways at once: each will be a new discovery and logic would not be foreseeable. To avoid this Wittgenstein supposes that they are all instances of a common form:

> In the sense in which there is a hierarchy of propositions there is, of course, also a hierarchy of truths and of negations, etc.
>
> But in the sense in which there are, in the most general sense, such things as propositions, there is only one truth and one negation.
>
> The latter sense is obtained from the former by conceiving the proposition in general as the result of the single operation which produces all propositions from the first level. Etc.
>
> The lowest level and the operation can stand for the whole hierarchy. [Wittgenstein, 1979, 7 Jan. 1917]

But if this is not to reintroduce the paradoxes, it seems that the operation that generates the whole hierarchy of propositions must be inexpressible. This is a blatant fudge, however. Either 'N' symbolizes in the same way each time or it does not. If it does, there is only one type of proposition and the paradoxes recur; if it does not, I am not a single finite intelligence, since I am capable of reading one sign in infinitely many different ways.

5.4 Ramsey's simple theory

The difficulty was to explain how an operation — something you can do to a proposition to yield another proposition — can be repeatable. If propositions are stratified into orders, so that applying an operation raises the order of the proposition, the worry is that it might be a different operation that is needed for the next application. Wittgenstein was convinced that there must be a sense in which a single operation could be repeated. Some time earlier he had referred to 'an operation whose own result can be taken as its base' [Wittgenstein, 1979, 23 Apr. 1916]. This notion of a single operation being repeatable was essential not only to guarantee a single general form of proposition, but also in order to be able to speak about the hierarchy of propositions.

> In this way, and in this way alone, is it possible to proceed from one type to another.
>
> And we can say that all types stand in hierarchies.
>
> And the hierarchy is only possible by being built up by means of operations. [Wittgenstein, 1979, 26 Apr. 1916]

What Wittgenstein seems not to have seen by the time he finished the last of his surviving notebooks in January 1917 was what this repeatable notion of operation was. The difficulty was that if operations were conceived as similar to functions, the vicious circle principle would ensure that they are not applicable to their own

results. If, for instance, the operation of negation is thought of as a function which takes a proposition p as argument and gives the proposition $\sim p$ as result, it presupposes a range of propositions which can serve as arguments to it and by the vicious circle principle the result of applying the function cannot be in the original range of arguments. It follows from this that in the proposition $\sim\sim p$ the two occurrences of the sign '$\sim$' are different symbols: in a logically perfect language the proposition would be written '$\sim_2\sim_1 p$ to indicate the difference between the two symbols.

The solution to this problem lies in Wittgenstein's notion of the sense of a proposition, which consists in the agreement or disagreement the proposition expresses with the truth-possibilities of elementary propositions. Although an operation will typically be specified by means of a transformation of signs (such as that involved in placing a '$\sim$' in front of any propositional sign 'p' to obtain the sign '$\sim p$'), the proposal is that we should think of the operation as individuated at the level not of signs but of senses. The reason is that the structure of senses is wholly different from that of signs and the paradoxes do not arise for them. This obviates the need to split senses into orders.

The threat of contradiction from the 'heterological' paradox remains, however. Ramsey, who gave a detailed discussion of this account, used the device of a hierarchy of meanings of 'meaning' to avoid it: the way in which a first order sign symbolizes is not the same as that in which a second order sign does. Suppose that we write 'means$_0$', 'means$_1$', etc. for the various meanings of 'means'. We can then define $\mathrm{Het}_n\widehat{z}$ to mean

$$(\exists\phi)(x \text{ means}_{n-1}\ \phi\widehat{z}\cdot \sim \phi x),$$

but the sense in which $\mathrm{Het}_n\widehat{z}$ has meaning cannot be the same as that in which 'means' is used in the definition. This blocks the paradox since the most we can say about this meaning is that

$$\text{'Het}_n\text{' means}_n\ \mathrm{Het}_n\widehat{z},$$

which is not sufficient to deduce the contradiction.

What Ramsey had done is to restrict the scope of formal logic to extensional functions by fiat, so that the need for the ramification in the formal part of logic is removed. But it does not disappear: the ramification is merely relocated to the metalanguage and emerges there as a hierarchy of meanings of 'meaning'.

5.5 Strong inexpressibility

Ramsey's solution is in fact a variant of Russell's proposal in the Introduction to the *Tractatus* of a hierarchy of ever more inclusive metalanguages. The difficulties caused by inexpressibility suggested to Russell

> that every language has, as Mr Wittgenstein says, a structure concerning which, *in the language*, nothing can be said, but that there may be

> another language dealing with the structure of the first language, and having itself a new structure, and that to this hierarchy of languages there may be no limit. Mr Wittgenstein would of course reply that his whole theory is applicable unchanged to the totality of such languages [i.e. languages I understand]. The only resort would be to deny that there is any such totality. [Wittgenstein, 1961, p. xxii]

To regard meaning$_0$, meaning$_1$, etc. as all being meanings of 'meaning' is to concede that there is one concept, namely *meaning*, of which they are all special cases, which is to dissolve the distinction Ramsey needed in order to block the intensional paradoxes. If, on the other hand, we insist that meaning$_0$, meaning$_1$, etc. are genuinely distinct concepts with only a verbal similarity to unite them, then I seem to be speaking not one language but infinitely many.

The dilemma Russell is briefly considering here — whether to regard the hierarchy of metalanguages as having a union — corresponds rather closely to Wittgenstein's dilemma considered earlier over whether to regard the series of N-operations as instances of a single operation. To see why, we need to bear in mind Wittgenstein's injunction to Russell that 'whatever a sign may express it can express'. Our formal language L_0 does not have the resources to express its own semantics, but that does not prevent us from conceiving of a more extensive language that does. In that more extensive language L_1 we can define a propositional function $\mathrm{Het}_1\widehat{z}$. In a further extension L_2 we can define $\mathrm{Het}_2\widehat{z}$. And so on. Since Het_n is not part of L_{n-1}, its existence does not threaten a formal contradiction in L_{n-1}. But let us now with intentional ambiguity use the sign Het in the language L_1 to mean the function $\mathrm{Het}_1\widehat{z}$. We know that when L_1 is embedded in L_2 there is a function $\mathrm{Het}_2\widehat{z}$ which the sign Het could also be taken to mean. But if the sign Het is capable of meaning the same as $\mathrm{Het}_2\widehat{z}$ when embedded in L_2 it is already capable of doing so in L_1. We cannot restrict what a sign means by fiat. The fact that the sign Het can be taken to mean $\mathrm{Het}_2\widehat{z}$ when it is embedded in L_2 shows that it has the right shape for the task of representing this function in L_2. But in that case it must have the right shape to represent this function in L_1. What a sign is capable of meaning is not a matter of what it can be defined to mean. We give signs meanings by reading them, and the only constraint on that is precisely what is unsayable, namely that the structure of the signs in question corresponds appropriately to the possibilities that exist in the world.

What this shows is that Ramsey's strategy of shifting intensional predicates to the metalanguage does not remove the difficulty we faced earlier over reconciling the ramified theory of types with the unity of the self: it merely relocates it. The only way of avoiding the ramification completely is not to take the very first step in its construction. We must insist that there is no meaning whatever of 'meaning' that allows us fully to say what 'Het' means. What we have arrived at is thus unmistakably the strong reading of the inexpressibility thesis, not the weak one: if we are to avoid the paradoxes and retain the unity of the self, we must insist that the relationship between a language and the world is something inexpressible not merely in that language but in any language.

It would be implausible to claim that the argument just given was Wittgenstein's own route to the inexpressibility thesis. He formulated that thesis quite early on — it occurs at the beginning of the notes dictated to Moore in April 1914,[Wittgenstein, 1979, p. 108] for instance — whereas the argument given here depends on a consequence of Wittgenstein's notion of sense that he himself seems not to have drawn until quite late on. The reason I have nonetheless mentioned the argument here is that it illuminates rather more clearly than Wittgenstein's own route to the conclusion the connection it has with the theory of types.

5.6 A Kantian theme

This is perhaps the point at which to mention an analogy between Wittgenstein's account and Kant's. The connexion lies in the observation that the general form of proposition plays the same role for Wittgenstein that the pure forms of intuition — space and time — play for Kant. According to Wittgenstein, given what objects there are for us to speak of, the structure of the general form of proposition determines what possibilities there are; if we could say what this range of possibilities is, we would have said what the limits are of what we can say. Correspondingly for Kant, given what objects-in-themselves there are for us to experience, the structure of space and time determines what the range of possible experience is; but we cannot coherently think through the limits of experience, for that is just what the resources we obtain from our sensibility are inadequate to ground. So the Tractarian difficulty of attempting to express the relationship between a sign and the world corresponds to the Kantian difficulty of thought about objects-in-themselves. Moreover, the difficulties in question are in both cases *partially* surmountable: although we cannot express the relationship between a sign and the world fully in a proposition, we can do so partially in a pseudo-proposition (which fails to be a proposition because it does not — and, as we have shown, cannot — participate fully in the logical structure of the language); this corresponds to Kant's claim that we can understand *noumena* only in a negative sense. Kant requires a transcendental deduction to show that the pure concepts of the understanding are applicable to experience, and that deduction takes as its central premise the unity of consciousness. For Wittgenstein, on the other hand, what would have to be transcendentally deduced is the relationship between language and the world. The deduction, which would once again take as its central premise the unity of the subject that is using the language, would leave us in equilibrium, using language to configure the world but unable fully to express that relationship.

What I have said about this last comparison between Wittgenstein and Kant does no more than suggest a loose analogy (although in truth it goes rather deeper than I have been able to indicate here), and hence a serious tension in Wittgenstein's account. The comparisons with Ramsey and Russell, on the other hand, allow us to see quite plainly the difference between their views of logic. The problem which the paradoxes pose for logic is that of preventing it from fragmenting into an infinity of parts with nothing intrinsic to unite them. Russell, at least in

the period leading up to *Principia Mathematica*, persistently sought the source of unification in an appropriate conception of the logical world. Ramsey, by contrast, sought it in our grasp of the metalanguage, but seems merely to have replicated the problem at that level. Wittgenstein solved the problem by the brutal method of banning at the first stage the sort of semantic reflection that is liable to generate it.

5.7 Nonsense again

We have seen, then, that Wittgenstein asserted not merely the weak thesis that there are features of my language which cannot be expressed within it, but the much stronger thesis that these features are such as to be inexpressible in *any* language whatever. They are, that is to say, literally unsayable. It has recently been popular among some commentators to characterize this view that 'nonsense is nonsense' as a *resolute* reading of the *Tractatus* in opposition to other, presumably less resolute readings. But we have some work still to do if we are to understand what this amounts to. If resolution amounted to no more than opposition to the Russellian view that what is unsayable in one language may be expressed in a suitable metalanguage, it would scarcely deserve the title of a 'reading' since it has been the interpretative orthodoxy for many years. (Even Russell himself explicitly regarded his view as an *opposition* to the main conclusion of the *Tractatus*. 'I do not believe his main thesis,' he wrote; 'I escape from it by a hierarchy of languages' [Grattan-Guinness, 2000, p. 440]). It is clear, moreover, that Wittgenstein intended his thesis to have far-reaching conclusions. If nonsense is simply nonsense, it follows, for instance, that logic in the ordinary sense is not applicable to it. We cannot, that is to say, unreflectively apply to nonsense the canons of reasoning that we normally apply to sense.

At the other extreme, though, it would be merely crass if the resolute view were to claim that attempts to say what can only be shown are indistinguishable from gibberish. This is plainly false as a matter of observed psychology: we evidently treat nonsense such as ' "*a*" refers to *a*' differently from nonsense such as 'gfsdlhkjfsdh' (even after we have read the *Tractatus*).

So if there is to be a genuine debate between resolute and irresolute readings of the *Tractatus*, it must avoid these two uninteresting extremes. The interest of such a debate lies in the importance of the sorts of statements that are in question: as we have seen, Wittgenstein included in the nonsensical not only semantic claims about the reference of words but arithmetical equations, the whole of ethics and aesthetics, and the nature of the self.

Now of course the observed fact that we do distinguish between important and unimportant nonsense is not itself philosophically significant in the present context: it might, after all, be a human failing to be persistently seduced by the misleading analogy between the nonsensical case and other formations that do make sense, but that would not show that we were *right* to do so. But Wittgenstein himself not only did not regard such things as ethics as gibberish but was not advocating

that we should try. Indeed the world's very significance to us is, as we have seen, one more instance of nonsense. So to treat all nonsense as gibberish would be to render the world insignificant. This would be radically impossible, since someone who succeeded in doing it would cease to be a willing subject at all.

6 THE THEORY OF CLASSES

6.1 The independence of elementary propositions

If the objects a and b have forms that permit them to be configured in the state of affairs ab, there is according to Wittgenstein nothing else — no 'propositional glue' — that is required in order that they *should* be so configured. This, I said earlier, is the nub of Wittgenstein's objection to Russell's theory of judgement. But if nothing else is *required* for ab to occur, nothing else can *prevent* it either: whether ab occurs is independent of whether any other state of affairs occurs.

So, at any rate, Wittgenstein believed. And it is here that the difference between his and Russell's versions of atomism is at its most pronounced. For Russell, the analysis of propositions to exhibit their underlying forms is in the end a practical project, subject at every stage to the limits of our current knowledge. It is therefore not absurd that there might be other aspects of the world, not encoded in our current analysis of propositions into atoms (and perhaps not directly knowable by us), that constrain the possibilities of combinations of objects to form states of affairs. Wittgenstein, on the other hand, simply assumed that there must be a complete analysis: for possibilities of occurrence of states of affairs in that analysis to be dependent on one another would thus be for there to be another aspect of the world not represented by those possibilities; and this would be for the analysis in question *not* to be in fact complete.

Wittgenstein accepted that it is impossible for a point in the visual field to be both red and green. But whereas this would have been for Russell evidence that elementary propositions need not be independent, Wittgenstein merely took it to show that 'This is red' and 'This is green' are not elementary. [Wittgenstein, 1961, 6.3751] He supposed, in other words, that they can be analysed somehow as truth-functional combinations of other, logically more basic propositions in such a way that their conjunction emerges as a *logical* contradiction. But Wittgenstein's careful lack of interest in the practicalities of this project led him not to bother with the question of how the analysis is actually done.

Notice, though, that Wittgenstein's argument here is wholly negative. No element in the system as so far described seemed to him to do the job of limiting the possibilities of occurrence of states of affairs. Moreover, the introduction of such an element would have wrecked Wittgenstein's account of logic by forcing there to be propositions formable using the N operator which express a priori impossibilities even though the way in which they are formed is incapable of revealing this. Wittgenstein would, in short, have had Kant's problem of the synthetic a priori to contend with. But Wittgenstein had little more than wishful thinking to give

as an argument against this.

It is therefore unsurprising that the doctrine of independence became the first part of the work to be publicly renounced by its author, perhaps because of his exposure in the 1920s to the Vienna Circle, whose members were interested not merely in abstract theory but in the details of its application. The difficulty lay in the incapacity of elementary propositions to attribute degrees of qualities. It is plain that 'X shines at a brightness of 1 lux' is incompatible with 'X shines at 2 lux'; but how are we to analyse these to display their relation? If the first is $E(b)$, we cannot, for example, analyse the second proposition as $E(b)\&E(b)$, since that would say no more than $E(b)$; and if we analyse it as $E(b')\&E(b'')$, the problem will recur when we try to explain the relationship between b' and b'' which makes them both instances of the same unit b, see [Wittgenstein, 1929].

6.2 Identity

There cannot be a genuine relation of identity between objects. For Wittgenstein this is just part of what it is to conceive of objects as simple. For if there were a genuine proposition saying that two objects are distinct, it would be possible for it to be false, i.e. for the two objects to be equal, in which case one at least of them would have to be a different object from the one it is.

But if identities are not genuine propositions with sense, that is not yet to say whether they are nonsense or senseless. So let us consider for a moment the idea that they are merely senseless, i.e. that '$a = a$' and '$a \neq b$' are tautologies. What is wrong with this view emerges quickly if we try to imagine them false. If we try to imagine a tautology being false, we find ourselves trying to imagine the simultaneous existence and non-existence of the same state of affairs, whereas the impossibility of $a \neq a$ is plainly of a different kind: no *state of affairs* is involved at all.

So is there *anything* that we could mean by '$a = a$'? Well, we might mean by this that a is the same object as a. But notice that when we regard this as a string of *signs*, it is not yet clear that what it attempts to say is even correct: if the first and the second instances of the sign 'a' referred to different things, it would be wrong to say that a is the same object as a. So we are in the same position with '$a = a$' as we were earlier with ' "p" says that p': when they are read as signs, they do not yet express anything, but when read as symbols they are correct without expressing anything significant.

If there is no genuine relation of identity that can hold between objects, we plainly ought not have any sign in our logical language purporting to express such a relation. Wittgenstein sketches how it is possible to do standard predicate logic without such a sign by adopting the convention that different names denote different objects. Moreover he proposes a corresponding convention for variables occurring within the scope of the same quantifier. To see how this operates in practice, consider the standard Russellian unpacking of a definite description. In

Russell's notation $\phi(\text{the } f)$ becomes some such expression as

$$(\exists x) : \phi x \cdot fx \cdot fy \supset_y y = x.$$

But in Wittgenstein's notation (which I shall write with a prime after the quantifier to distinguish it from Russell's) it becomes

$$(\exists x)' : \phi x \cdot fx \cdot \sim (\exists x, y)' : fx \cdot fy.$$

Wittgenstein's notation is extremely tedious to use in practice, but that is scarcely the point. What matters is that the formulae which *cannot* be expressed in Wittgenstein's notation, such as the Russellian $(\exists x).x = a$, turn out to be things that we don't want to be able to express anyway, because they do not really say anything significant.

But at this point there is a difficulty. Russell in *Principia Mathematica* had simply *defined* identity by treating $x = y$ as an abbreviation for the claim that x and y have all their properties in common. Wittgenstein evidently had to find an error in this definition if he was to maintain that there is no such relation as identity.

What is wrong with Russell's definition is that there are possible worlds in which it incorrectly makes distinct objects come out equal. If we have two distinct objects a and b of the same form, it is at least possible (by the independence of elementary propositions) that they should have all their elementary properties in common, i.e. that whenever an elementary proposition $\phi(a)$ is true, the corresponding proposition $\phi(b)$ is also true and vice versa. But if this holds for elementary properties, it will hold for *all* properties, because Wittgenstein's logic is truth-functional. So in that case a and b will have all the same properties, and will therefore come out as equal by Russell's definition, which by hypothesis they are not.

It is worth noting, though, that the possible world in which a and b have all the same external properties cannot be the actual world. If two things have all their properties in common,

> it is quite impossible to indicate one of them. For if there is nothing to distinguish a thing, I cannot distinguish it, since otherwise it would be distinguished after all. [Wittgenstein, 1961, 2.02331]

In other words, we have here a feature which distinguishes the actual world from other possible worlds. A world in which two things have all the same properties, although possible, could not be the actual world. Or rather, if it were the actual world, we would represent it using a different language containing fewer names.

6.3 Classes

We have seen that according to Wittgenstein there is not a genuine relation of identity between objects. But in that case there can be no genuine theory of classes of the sort required by a Russellian reduction of mathematics. To see why,

let us recall that Russell had in *Principia* adopted a 'no-class' theory according to which class terms are incomplete symbols which should, strictly speaking, be eliminated from contexts where they occur, in much the same manner as definite descriptions. So a statement which appears to be about the class $\{x : \phi x\}$ is rewritten so as to be about the propositional function $\phi\widehat{x}$. I shall call this an *accidental* class term: that is to say, we are to conceive of the class $\{x : \phi(x)\}$ as having as members in any world just those objects which satisfy ϕ in that world. But consider now '$\{a, b\}$'; I shall call this class term *essential*, by which I mean that it denotes a class whose members in any world are the same, no matter what properties those objects have in that world. But how are we to talk about this class? We should like to rewrite a sentence involving $\{a, b\}$ so as to be about the function $\widehat{x} = a \vee \widehat{x} = b$ instead; but according to Wittgenstein we cannot, because there is no such propositional function.

We have two types of class terms, then, accidental and essential. The difference between them lies in how they track objects across possible worlds: the accidental class has as its members in each world the objects which have in that world the class's defining property; the essential class has the same members in each world, no matter what properties those objects have in any world. Wittgenstein's logic allows us to talk about classes only in the accidental idiom, not the essential one.

But it is the essential idiom that we need if we are to achieve a Russellian reduction of mathematics. If, for instance, we define the number *one* to be the class of all single-element classes, this will work only on the essential understanding of the notion. For if we try to use the accidental notion, we face the difficulty that there is a possible world in which every object shares its properties with some other object, in which case there are no one-membered classes and the number *one* comes out as empty. This has the awkward consequence that simple numerical inequalities such as $1 \neq 2$ come out as true only accidentally Hence Wittgenstein's remark:

> The theory of classes is altogether superfluous in mathematics.
>
> This is connected with the fact that the generality required in mathematics is not *accidental* generality. [Wittgenstein, 1961, 6.031]

What Wittgenstein means is that the only sort of theory of classes that his logic can generate is the accidental one, and any attempt to base mathematics on *that* would make it hostage to accidents, whereas Wittgenstein takes it as evident that mathematics must be independent of how the world happens to be.[4]

6.4 The number of things

How many objects are there of a particular type? It is clear that it cannot be a significant proposition in Wittgenstein's system to say that there are n of them, because if $a_1, a_2, \ldots, a_n$ are indeed all the objects of this type, then this could not

[4] Added weight is lent to the interpretation just outlined by the fact that Wittgenstein wrote by this remark in Ramsey's copy of the *Tractatus* the word 'Identity'.

be otherwise. But — as in other such cases we have met before — we have further work to do to determine whether it is in that case a tautology that there are n objects of this type, or nonsense.

So let $T\widehat{x}$ be a function that is true of everything of the type in question, e.g. the function $f\widehat{x} \vee \sim f\widehat{x}$. Let q_n be the sentence

$$(\exists x_1)'(\exists x_2)' \ldots (\exists x_n)' : Tx_1 \cdot Tx_2 \ldots Tx_n \tag{1}$$

in Wittgenstein's notation. It might seem at first sight as if q_n says that there are at least n different objects satisfying the propositional function $T\widehat{x}$, i.e. that there are at least n objects of the type in question. If this were right, then of course $q_n \cdot \sim q_{n+1}$ would say that there are exactly n objects of the type. This would then be either a tautology or a contradiction depending on the value of n.

But this is wrong. What is wrong with it is that although on Wittgenstein's account (1) is indeed true if there are at least n objects of this type, it is not false but meaningless if there are fewer than n. To see why, consider for simplicity the case where there are just three objects a_1, a_2, a_3 of the given type. Then q_3 expresses the tautology

$$Ta_1 \cdot Ta_2 \cdot Ta_3$$

as desired. But what does q_4 express? Our only resources at this point are the three tautologies Ta_1, Ta_2, Ta_3, and it is clear that no combination of them that expresses contradiction could reasonably be seen as expressed by

$$(\exists x_1)'(\exists x_2)'(\exists x_3)'(\exists x_4)' : Tx_1 \cdot Tx_2 \cdot Tx_3 \cdot Tx_4.$$

Whether an existential quantifier makes sense when interpreted according to Wittgenstein's symbolic convention thus depends on how many names there are in the language that could be substituted for the variable of quantification. Confirmation that this is the correct interpretation of Wittgenstein's intentions is supplied by one of the remarks he wrote in the margin of Ramsey's copy of the *Tractatus* in 1923:

> The prop[osition] 'there are n things such that ...' presupposes for its significance what we try to assert by saying 'there are n things'. [Lewy, 1967, 421]

By extension the same thing applies to the axiom of infinity, which tries to say that there are $\aleph_0$ things: this is significant only if it is true. [Wittgenstein, 1961, 4.1272]

But if that is the case, we might think that the axiom of infinity admits a transcendental proof as follows. My language — '*the* language which I understand' [Wittgenstein, 1922, corrected edn., 5.62] — is fashioned with the primary purpose of saying how the world is and, correlatively, how it is not. And whether or not there are in fact infinitely many things, we can represent in language the *possibility* that there are. But a language in which there were only finitely many names would

not be able to represent this possibility. Hence there must be infinitely many names in our language.

Ramsey wrote down something like this argument shortly after his visit to Wittgenstein:

> I can say 'There are an infinite number of atoms'. This may be false, but it is possibly true and therefore significant. And for it to mean anything there must be an infinite number of things. These need not be material objects because the proposition requires analysis to discover its real form .... But roughly it involves the possibility of saying 'Here is an atom' about an infinity of 'here' 's. [Ramsey, 1991, 175]

Ramsey never published this argument, presumably because he very soon afterwards adopted a different theory of identity according to which the axiom of infinity is either a tautology or a contradiction but always significant, and on this view the argument plainly does not work. But Wittgenstein strongly disagreed with Ramsey's account of identity (see below), so he never had Ramsey's reason not to pursue this transcendental argument for the axiom of infinity. Whether the argument would have appealed to him on its own merits is harder to judge. Later, when he returned to philosophy at the end of the 1920s, he began to explore the idea that our talk of the infinite can be analysed in such a way that it is merely part of the way we describe the finite and does not require there to be infinitely many objects: on that conception the argument once again fails, but there does not seem to be any trace of this thought in the *Tractatus*. [5]

6.5 Identity and propositional functions in extension

I have said in Section 6.2 that according to Wittgenstein $a = a$ and $a = b$ are always nonsense if a and b are Tractarian objects. What we did not examine then was the idea of simply defining $a = a$ to be a tautology and $a = b$ to be a contradiction. More formally, suppose that we define a function Q of two objectual variables by the prescription

$$Q(x, y) = \begin{cases} \text{Taut} & \text{if } x = y \\ \text{Cont} & \text{otherwise} \end{cases}$$

The idea is that the function Q is to take the place of identity in contexts where it is required.

Thus, for instance, we immediately retrieve the essential notion of class by replacing talk of any class A by talk of the logical sum of the functions $Q(\widehat{x}, a)$ for $a \in A$. This is in effect what Ramsey proposed to do in his 1925 paper 'The foundations of mathematics'. (I say 'in effect' because he in fact adopted an indirect procedure which makes his definition seem a little less circular than the

[5] For more on Ramsey's transcendental argument, see [Potter, 2005].

one I have just given.) Ramsey thus claimed to show that the theory of classes (essential classes) consists of tautologies in Wittgenstein's sense.

As Ramsey observed straightaway, [Ramsey, 1990, p. 216] there are two important issues to address concerning his proposal to introduce this new notation: is it consistent? and is it legitimate? In his original article Ramsey correctly points out that the answer to the first question is yes: introducing a new symbol for tautology or contradiction cannot create new senses, but will merely provide us with further ways of expressing old ones; it therefore cannot result in the sort of explosion of the domain of senses that leads to paradox.

On the other hand, to the second question — whether the notation is legitimate — Ramsey answers in the affirmative without giving any more reason for his answer than the bald statement that 'it is an intelligible notation, giving a definite meaning to the symbols in which it is employed'. [Ramsey, 1990, p. 216] Wittgenstein's first reaction to Ramsey's procedure, in a letter he wrote in 1927, was to explain why Ramsey's function Q, although 'a very interesting function', could not be a definition of identity. Ramsey's response to this was to say in effect that he and Wittgenstein were talking past one another. Ramsey had not claimed that Q *was* the identity function, since he agreed with Wittgenstein's argument that there could be no such function. All Ramsey claimed was that the function Q could be used in quantified contexts to express correctly various senses that one wants to be able to express. Take, for example, 'something other than a is f'. This can be expressed using Wittgenstein's convention for variables as

$$fa. \supset .(\exists x, y)' : fx \cdot fy \cdot :\sim fa. \supset .(\exists x)'.fx.$$

But using the more usual Russellian convention for quantifiers together with Ramsey's Q function we can also express it as

$$(\exists x).fx \cdot \sim Q(x, a).$$

As Ramsey says in his letter to Wittgenstein, 'if you admit $Q(x, y)$ as a legitimate symbol at all it seems to me this must be all right'. [Wittgenstein, 1995, p. 220]

It was not until some time later, after Ramsey's death, that Wittgenstein managed to formulate a clear explanation of what was *really* wrong with Ramsey's notation. We can of course introduce the symbol $Q(a, a)$ to mean tautology and the symbol $Q(a, b)$ to mean contradiction, but by doing so we do not license the replacement of a with a variable to form a function $Q(\widehat{x}, a)$. The symbol $Q(a, a)$, as we have now introduced it, is a *word* and does not have the sort of complexity that would entitle us to decompose it for quantificational purposes. Treating identity, so defined, as a function would be like treating 'coal', 'coat' and 'corn' as three values co(al), co(at), co(rn) of a single function 'co' [Wittgenstein, 1974, II, Section 16].

6.6 Mathematics as tautological

That is clear enough, and plainly right as far as it goes, but it does not go very far: Wittgenstein's objection applies to Ramsey's specific proposal, sure enough, but

it does not seem to provide us with a *general* reason to reject the idea that mathematics consists of tautologies. At this stage in the debate, then, it seems correct to conclude, as Sullivan did in his discussion of the correspondence,[Sullivan, 1995] that the exchange between Wittgenstein and Ramsey is disappointingly localized, a narrow matter of the correctness of a technical device which no-one any longer proposes to use. Yet it seems clear from the fact that Wittgenstein returned to the matter repeatedly, and evidently expended considerable effort on puzzling out just what was wrong with Ramsey's proposal, that Wittgenstein saw it as a significant threat. If so, what Wittgenstein found threatening can only have been the general claim that mathematics consists of tautologies. What, then, were Wittgenstein's general reasons, independent of the debate with Ramsey, for rejecting *that*?

What is striking is that Wittgenstein should have been so opposed to this idea when he went out of his way to emphasize in the *Tractatus* how *similar* the equations of mathematics are to tautologies:

> The logic of the world, which is shown in tautologies by the propositions of logic is shown in equations by mathematics. [Wittgenstein, 1922, 6.22]

If this is the similarity, then, what is the difference? The fundamental difference between tautologies and equations lies in how they can be applied. A tautology such as $p \vee \sim p$ can be seen as a sort of limiting case of a genuine proposition. Its component parts, such as p, have sense, and the ways in which those parts are combined to form the whole are ways in which propositions with sense can be formed. It is just that in this case the sense so formed turns out to be empty. What happens when we try to form a parallel explanation of the equation $7 + 5 = 12$? The intended application is that this equation allows us to infer such facts as that if there are 7 apples and 5 oranges then there are 12 pieces of fruit. In symbols,

$$\mathrm{N}\phi = 7 \cdot \mathrm{N}\psi = 5 \cdot \mathrm{N}x(\phi x \cdot \psi x) = 0 \supset \mathrm{N}x(\phi x \vee \psi x) = 12. \qquad (2)$$

But this is now plainly not parallel to the tautology case. For no one instance of (2) carries the import of the equation $7+5 = 12$. And if we try to treat the equation as meaning the universal generalization of (2), we run into the difficulty that we can only generalize over one level in the hierarchy at a time, which is not what we want: we ought to be able to count first-level propositional functions by just the same means as we count apples. But even if we prescind from this difficulty and focus only on the case where what we are trying to count are Wittgensteinian objects, we still do not get what we want: it is possible that no first-level property has just five instances, and in that case the equation $7 + 5 = 13$, interpreted according to the current proposal, would come out true, as would every other equation with the number 5 on the left hand side (because the antecedent of the conditional would be uninstantiated). This sort of accidental truth is plainly not what we were aiming for, so the only thing left to us is to interpret $7 + 5 = 12$ as meaning that (2) is not merely always *true* but always *tautological*. This, though, plainly cannot *itself* be a tautology since it is at the wrong semantic level for that.

If we wish to claim that mathematics consists of tautologies, then, it is no use treating equations as merely abbreviated embodiments of their intended applications: the only course is to treat them as tautologies in their own right, their tautologousness not being seen as derived from their applications. Ramsey's account is evidently an instance of this general strategy. But if we do this, we shall eventually have to explain how these tautologies nevertheless do get applied: we shall have to establish a connection between the new ways of expressing senses thus introduced and the old ones. But now our difficulty is that we have broken the link with a crucial aspect of Wittgenstein's account of tautologies, namely that they can be seen as limiting cases of genuine propositions, i.e. as trivial cases of forms capable of expressing non-trivial senses. Without that link mathematics floats free of the rest of language and the account lapses into a version of formalism.

7 ARITHMETIC

7.1 Numbers

We have said what mathematics is not — tautologies. Now we turn to what it is. The only part of mathematics which Wittgenstein deals with explicitly is elementary arithmetic, and although he does not give any explanation of this restriction, he seems to have assumed that this was the only part of mathematics which makes contact with reality — the only part, that is to say, which is directly applicable.

The idea Wittgenstein fell on to explain the applicability of arithmetic takes as its inspiration a treatment of numbers in terms of numerically definite quantifiers that Frege had considered (but in the end rejected) in the *Grundlagen*. That treatment, when rewritten in Wittgenstein's notation, involves defining

$$
\begin{aligned}
(\exists^{\geq 1}x).fx &=_{\mathrm{Df}} (\exists x)'.fx \\
(\exists^{\geq 2}x).fx &=_{\mathrm{Df}} (\exists x, y)'.fx \cdot fy \\
&\vdots \\
(\exists^{\geq n+1}x).fx &=_{\mathrm{Df}} (\exists x)'(\exists^{\geq n}y).fx \cdot fy \\
&\vdots
\end{aligned}
$$

If we define

$$(\exists^{n}x).fx =_{\mathrm{Df}} (\exists^{\geq n}x).fx \cdot \sim (\exists^{\geq n+1}x).fx,$$

it is then straightforward to derive an account of the use of numerals in counting things.

But for Wittgenstein's purposes this account is too specific: it cannot explain how numbers can also be used to count not just things but the terms of other

formal series, such as the one we discussed earlier.

$$\begin{aligned} aR^1b &=_{\mathrm{Df}} aRb \\ aR^2b &=_{\mathrm{Df}} (\exists x) : aRx \cdot xRb \\ &\vdots \end{aligned}$$

In order to achieve this Wittgenstein generalized Frege's treatment so that numerals became now indices of the repetition of *any* operation whatever. If Ω is any operation and p is any proposition, Wittgenstein lets

$$\begin{aligned} \Omega^0(p) &=_{\mathrm{Df}} p \\ \Omega^{\nu+1}(p) &=_{\mathrm{Df}} \Omega(\Omega^\nu(p)) \end{aligned}$$

The numerals $0, 0+1, 0+1+1, \ldots$ (which are of course normally abbreviated $0, 1, 2, \ldots$) thus act, when occurring in exponent position, to count the number of times the operation Ω has been applied to the proposition p.

7.2 Equations

In order that we should now be able to develop arithmetic, we might expect definitions of the arithmetical functions by recursion, e.g.

$$\begin{aligned} \Omega^{\mu+0}(p) &=_{\mathrm{Df}} \Omega^\mu(p) \\ \Omega^{\mu+(\nu+1)}(p) &=_{\mathrm{Df}} \Omega^{\mu+\nu}(\Omega(p)). \end{aligned}$$

But instead Wittgenstein proposes to use more compact definitions of addition and multiplication that exploit the direct links between these functions and composition.

$$\begin{aligned} \Omega^{\mu+\nu}(p) &=_{\mathrm{Df}} \Omega^\mu(\Omega^\nu(p)) \\ \Omega^{\mu\nu}(p) &=_{\mathrm{Df}} (\Omega^\mu)^\nu(p) \end{aligned}$$

(At any rate, he states explicitly the definition of multiplication and leaves the reader to work out the corresponding one for addition.) Wittgenstein does not say explicitly whether he prefers these to the recursive definitions merely because they are more compact or whether he has some objection to the method of recursive definition. If the latter, it would of course make a significant difference to his system, since it is far from obvious how to give correspondingly compact definitions of other arithmetical functions such as exponentiation.

With the definitions of addition and multiplication just given we can prove

explicit numerical equations such as $2 + 2 = 4$ or $2 \times 2 = 4$. Thus, for example,

$$\begin{aligned}\Omega^{2+2}(p) &= \Omega^2(\Omega^2(p)) \\ &= \Omega(\Omega(\Omega^2(p))) \\ &= \Omega(\Omega^{2+1}(p)) \\ &= \Omega^{(2+1)+1}(p) \\ &= \Omega^{3+1}(p) \\ &= \Omega^4(p).\end{aligned}$$

The idea in general, then, is that an equation $\mu = \nu$ is to mean that $\Omega^\mu(p) = \Omega^\nu(p)$. Elsewhere in the *Tractatus* Wittgenstein indicates by the gnomic remark that '$\sim\sim\ p = p$' that the identity sign here is to mean identity of *sense*. So $\Omega^\mu(p) = \Omega^\nu(p)$ is to be interpreted as meaning that $\Omega^\mu(p)$ and $\Omega^\mu(p)$ have the same sense, or equivalently that $\Omega^\mu(p) \equiv \Omega^\nu(p)$ is a tautology.

Now for this to give an extensionally correct account of the equation $\mu = \nu$ we must plainly mean to be asserting this for *every* operation Ω. (If it was sufficient that it hold for *some* operation Ω, the case $\sim\sim\ p = p$ would show that $2 = 0$.) But we cannot achieve this effect by genuine quantification because that would involve quantifying into the expression

$$\text{that } \Omega^\mu(p) \equiv \Omega^\nu(p) \text{ is a tautology,}$$

which is not a significant proposition but nonsense. So if the account is to come out right, we shall presumably have to regard Ω and p as schematic. As predicted in the previous chapter, then, Wittgenstein's account treats each arithmetical equation as abbreviating not a tautology but a claim that all the symbols of a certain form *express* tautologies.

It is evident even from the layout of the sections of the *Tractatus* that Wittgenstein intended to stress the parallels between logic and arithmetic. What the apparent propositions of logic and arithmetic share is that they do not themselves express thoughts about the world, but are used to license inferences between genuine propositions that do express thoughts.

> Indeed in real life a mathematical proposition is never what we want. Rather, we make use of mathematical propositions *only* in inferences from propositions that do not belong to mathematics to others that likewise do not belong to mathematics. [Wittgenstein, 1961, 6.211]

But despite these similarities we can now see that there is also a fundamental difference between logic and arithmetic. A tautology and an arithmetical equation, although both can be used in inference, are of different semantic levels. A tautology is on the same level as the propositions with sense but fails to say anything significant only through its vacuity. An equation, on the other hand, tries to say that symbols of a certain form express tautologies. The difference, in other words,

is the same as that between the vacuous proposition $p \vee \sim p$, which expresses tautology, and the pseudo-proposition that $p \vee \sim p$ is a tautology, which tries to say what the former proposition shows.

7.3 Quantification

In the account we have given of the meaning of the equation '$\mu = \nu$' the letters 'μ' and 'ν' are schematic, standing not merely for arbitrary numerals (such as '4' or '126') but more generally for arbitrary number terms (such as '$7 + 5$' or '$4 \times (6 + 2)$'). This allows us to give a meaning to schematic general statements about numbers. For example, the schema $\mu + 1 = 1 + \mu$ means that

$$\Omega^{\mu+1}(p) = \Omega^{1+\mu}(p),$$

with μ, Ω, and p all schematic. Nothing similar can be done with existential generalizations about numbers, however. If, for instance, we try to represent $(\exists x).\mu(x) = 0$ as meaning that

$$(\exists x)(\Omega^{\mu(x)}(p) \equiv p \text{ is a tautology}),$$

we get the wrong meaning. To get the correct meaning we should have to write

$$(\exists x)(\Omega)(p)(\Omega^{\mu(x)}(p) \equiv p \text{ is a tautology}),$$

but this is meaningless for the by now familiar reason that the expression

$$\Omega^{\mu(x)}(p) \equiv p \text{ is a tautology}$$

is not a proposition but a pseudo-proposition and hence cannot meaningfully be quantified.

We can thus coherently express universal but not, in general, existential generalizations about numbers. To be able to express generalizations is not yet to be able to prove them, though. The proof that in general $\mu + 1 = 1 + \mu$, for instance, involves induction on μ, and nothing we have said so far licenses that. Indeed the *Tractatus* makes no explicit mention of the principle of mathematical induction at all. (The two references in the text to 'induction' seem to be to the law of induction in physics.)

At an even more basic level, Wittgenstein's definition of multiplication, for example, does not give us any direct assurance that the product of two natural numbers is a natural number, i.e. that if 'μ' and 'ν' are numerals there is a numeral 'σ' such that

$$\Omega^{\mu\times\nu}(p) = \Omega^{\sigma}(p)$$

with Ω and p arbitrary. When we calculate $\mu \times \nu$ in any particular case (such as Wittgenstein's calculation of '2×2'), we of course find that there is such a numeral (in this case '4'), but what *general* reason do we have for thinking that there is always one? The usual proof involves a double induction on μ and ν.

It may seem as though such appeals to induction would be a further step, requiring a new argument. On the purely formal level this is undoubtedly so, but not, I think, on a conceptual level. The reason is that in order to ask the question whether $\mu \times \nu$ is in general a number we have already to suppose that $\mu \times \nu$ has been defined. And the definition, whose coherence is not at this point in question, makes an implicit appeal to induction for its meaningfulness. This is because our understanding of the numerals 'μ' and 'ν' occurring in the definition is in turn inductive: numerals are, according to Wittgenstein's definition, symbols which can be obtained from zero inductively by adding one. We are tempted to deny this dependence because it seems that for any *particular* numerals 'μ' and 'ν' no appeal to induction is required: we merely apply the definition a finite number of times. But what this overlooks is that *each* time we apply the definition we have to understand it as the *same* multiplication function that is being calculated. And for this we do require the general grasp of the number series that is given to us by its inductive definition.

Wittgenstein's definition of number thus commits him to the validity of argument by induction on the numerals. That does not, however, explain how we ground its validity. Wittgenstein's answer to that lies in the remarks which head the section of the *Tractatus* on arithmetic and logic:

> The general form of a truth-function is $[\overline{p}, \overline{\xi}, \mathrm{N}(\overline{\xi})]$.
>
> This is the general form of a proposition.
>
> What this says is just that every proposition is a result of successive applications to elementary propositions of the operation $\mathrm{N}(\overline{\xi})$. [Wittgenstein, 1922, 6–6.001]

The point of making all the remarks on arithmetic subsidiary to this is that the general form of proposition is recursive: propositions are obtained from elementary propositions by iterations of the N operation. Since all operations apply to propositions, all operations are ultimately expressible in terms of the N operation. But numbers are merely a notation for expressing the repeated application of operations. So our understanding of numbers in general is also based on our understanding of the general form of proposition.

None of this, of course, amounts to an explanation of the validity of induction. What it does, though, is to base that on an appeal to something Wittgenstein takes to be fundamental, namely that grasp of language as a whole which we have when we understand the general form of proposition. Wittgenstein sees this as playing a central rôle in his theory corresponding to the rôle of intuition in Kant. Kant talks here of the 'limitlessness in the progression of intuition'; [Kant, 1781, A25] Wittgenstein needs the general form of proposition to do the same thing.

> The question whether intuition is needed for the solution of mathematical problems must be given the answer that in this case language itself provides the necessary intuition.

> The process of *calculation* serves to bring about that intuition. [Wittgenstein, 1922, 6.233]

7.4 Dependence on infinity

Does Wittgenstein's account rule out the possibility that the numbers are not all distinct? If 'μ' and 'ν' are distinct numerals, can we conclude, in other words that $\mu \neq \nu$? In order to prove the corresponding result in their system, Whitehead and Russell needed to assume an axiom of infinity asserting that there are infinitely many individuals, but the axiom of infinity cannot play the same rôle in Wittgenstein's account. The infinitude of the world (if it is indeed infinite) shows itself in the language we use to describe it and cannot be expressed by any one proposition in the language. However, there are two different ways in which the world might be infinite corresponding to the two aspects it has, that of permanence and that of possibility. For the world to be infinite in its permanent aspect there would have to be infinitely many objects, and this would be shown by there being infinitely many names in the language. For there to be infinitely many possibilities as to how the world could be, on the other hand, there would have to be infinitely many states of affairs in the world, and this would be shown by there being infinitely many elementary propositions in the language.

Now both these — the number of names and the number of elementary propositions — are ways in which language reflects the limits of empirical reality. 'Empirical reality is limited by the totality of objects. The limit also makes itself manifest in the totality of elementary propositions' [Wittgenstein, 1922, 5.5561]. But the connection between the two sorts of limit is even closer than this suggests: empirical reality is infinite in the first respect if and only if it is infinite in the second.

Suppose now for a moment that there is only a finite number of elementary propositions. Then there are only finitely many propositional senses, and consequently the set of all operations, i.e. functions from propositional senses to propositional senses, is finite too. A straightforward group-theoretic argument then shows that there must be numbers M, N with $M < N$ such that for every operation Ω and every proposition p we have $\Omega^M(p) = \Omega^N(p)$. But this is just what we said was meant by $M = N$. Consequently there are no natural numbers other than $0, 1, 2, \ldots, N-1$. By contraposition it follows that if there are infinitely many numbers then there are infinitely many elementary propositions, and hence infinitely many objects.

A further consequence is that subtraction cannot be defined in Wittgenstein's system. For if it could, we would have $N - M = 0$, and so $\Omega^{N-M}(p) = p$ for every operation Ω and proposition p; in particular, if we apply this in the case where $\Omega(p) = p_0$ for all p, where p_0 is some fixed proposition, we obtain $p = p_0$ for all p, i.e. all propositions have the same sense. Contradiction.

Although reality is limited by how many objects and elementary propositions there are, this is not a limit we can express in our language. The possibilities

we can express are possibilities as to the ways the objects could be configured, not possibilities as to how many objects there are. So if the form of arithmetic is dependent on how many objects there are, that does not directly contradict Wittgenstein's injunction that 'the generality that is required in mathematics is not *accidental* generality'. The reason the dependence of arithmetic on the number of objects nevertheless seems embarrassing to Wittgenstein is that there is plainly another sense in which it *is* possible that the number of objects should be different from what it is, namely that logic allows for a different number of objects even if language as a whole does not. Language draws the limits of the world, the limits of what is possible. But we can divide our grasp of language into two parts, the elementary propositions on the one hand and the general form of proposition on the other; and corresponding to this division we can detect two levels of possibility that language as a whole allows for. 'Mathematics', Wittgenstein says, 'is a logical method' [Wittgenstein, 1922, 6.2]. A consequence of this is that mathematics should not limit the world any more than logic does. If mathematics is to lie at the same level of generality as logic, it must be necessary not only with respect to the level of contingency that our language expresses, namely contingency as to which arrangements of objects into states of affairs actually occur, but also with respect to the higher level inexpressible in the language itself but represented by the general form of proposition, namely contingency as to what objects there are. So Wittgenstein does indeed seem to be in need of an argument to show that *our* language is one that represents there to be infinitely many objects.

ACKNOWLEDGEMENTS

Parts of this chater are based on notes of my lectures taken by Arif Ahmed. I am grateful to Peter Sullivan for comments on an earlier draft.

BIBLIOGRAPHY

[Anscombe, 1967] G. E. M. Anscombe. *An Introduction to Wittgenstein's Tractatus*, 3rd edn. (London: Hutchison, 1967).

[Geach, 1981] P. T. Geach. Wittgenstein's operator N. *Analysis*, 41 (1981), 168–71.

[Grattan-Guinness, 2000] I. Grattan-Guinness. *The Search for Mathematical Roots:Logics, Set Theories and the Foundations of Mathematics from Cantor through Russell to Gödel* (Princeton University Press, 2000).

[Kant, 1781] I. Kant. *Kritik der reinen Vernunft* (Riga: Hartknoch, 1781; B edn. 1878) (trans. as *Critique of Pure Reason*).

[Kant, 1929] I. Kant. *Critique of Pure Reason*, trans. by Norman Kemp Smith (London: Macmillan, 1919).

[Lewy, 1967] C. Lewy. A note on the text of the *Tractatus*, *Mind*, 76 (1967), 416–23.

[McGuinness, 1989] B. McGuinness. Wittgenstein's pre-*Tratatus* manuscripts. In *Wittgenstein in Focus — Im Brennpunkt: Wittgewnstein*, ed. by Brian McGuinness and Rudolf Haller, Grazer philosophische Studien (Amsterdam: Rodophi, 1989), pp. 35–47.

[Potter, 2005] M. Potter. Ramsey's transcendental argument. In *Ramsey's Legacy*, ed. by Hallvard Lillehammer and D. H. Mellor (Oxford University Press, 2005), pp. 71–82.

[Ramsey, 1991] F. P. Ramsey. *Notes on Philosophy, Probability and Mathematics*, ed. by Maria Carla Galavotti, History of Logic, 7 (Naples: Bibliopolis, 1991).

[Ramsey, 1990] F. P. Ramsey. *Philosophical Papers*, ed. by D. H. Mellor (Cambridge University Press, 1990).
[Russell, 1910] B. Russell. *Philosophical Essays* (London: Longman Green, 1910).
[Sullivan, 1995] P. M. Sullivan. Wittgenstein on *The Foundations of Mathematics*, June 1927, *Theoria*, 61 (1995), 105–42.
[Sundholm, 1992] G. Sundholm. The general form of the operation in Wittgenstein's *Tractatus*, in *Criss-crossing a Philosophical Landscape*, ed. by Joachim Schulte and Göran Sundholm, Grazer philosophische Studien (Amsterdam: Rodophi, 1992), pp. 57–76.
[von Wright, 1990] G. H. von Wright. *A Portrait of Wittgenstein as a Young Man: From the Diary of David Hume Pinsent 1912–1914* (Oxford: Blackwell, 1990).
[Wittgenstein, 1922] L. Wittgenstein. *Tractatus Logico-Philosophicus* (London: Kegan Paul a& Trubner, 1922; corrected edn. 1933).
[Wittgenstein, 1929] L. Wittgenstein. Some remarks on logical form. *Proc. Arist. Soc., Supp. Vol., 9* (1929), 162–71.
[Wittgenstein, 1961] L. Wittgenstein. *Tractatus Logico-Philosophicus*, trans. by D. F. Pears and B. F. McGuinness (London: Routledge & Kegan Paul, 1961).
[Wittgenstein, 1971] L. Wittgenstein. *Prototractatus: An Early Version of Tractatus Logico-Philosophicus*, ed. by B. F. McGuinness, T. Nyberg and G. H. von Wright, trnas. by D. F. Pears and B.F. McGuinness, intro. by G. H. von Wright (London: Routledge & Kegan paul, 1971).
[Wittgenstein, 1974] L. Wittgenstein. *Philosophical Grammar*, ed. by Rush Rees and trans. by Anthony Kenny (Oxford: Blackwell, 1974).
[Wittgenstein, 1975] L. Wittgenstein. *Philosophical Remarks*, ed. by Rush Rees and trans. by Raymond Hargeraves and Roger White (Oxford: Blackwell, 1975).
[Wittgenstein, 1979] L. Wittgenstein. *Notebooks, 1914–1916*, ed. by G. H. von Right and G. E. M. Anscombe, 2nd edn (Oxford: Blackwell, 19740.
[Wittgenstein, 1995] L. Wittgenstein. *Cambridge Letters: Correspondence with Russell, Keynes, Moore, Ramsey and Sraffa*, ed. by Brian McGuinness and G. H. von Wright (Oxford: Blackwell, 1995).

LEŚNIEWSKI'S LOGIC

Peter M. Simons

1 LIFE

Stanisław Leśniewski was born on 30 March 1886 at Serpukhov in Russia, where his father Izydor was an engineer engaged in building the Trans-Siberian railway. He attended school in Irkutsk, before studying in Western Europe (Leipzig, Zurich, Munich). He arrived at Lvov in Austrian Poland in 1910 to work on his Ph.D. under Kazimierz Twardowski. In the years before World War I he wrote several essays on the philosophy of language and logic. In 1911 he read Łukasiewicz's book *On the Principle of Contradiction in Aristotle* and in it he learnt about symbolic logic and Russell's Paradox. This set him on his life's path of producing an antinomy-free system of logical foundations for mathematics. He received his doctorate in 1911 and married the following year, before travelling in France, Italy and Russia. During World War I he lived and worked in Moscow, but in 1918 after the October Revolution he returned to Poland and after failing to gain his habilitation in Lvov, obtained it in Warsaw. During the 1919–21 Polish–Bolshevik War he worked as a codebreaker for the Polish General Staff's Cipher Bureau, contributing in no small way to the defeat of the Red Army. In 1919 he was appointed Professor of the Foundations of Mathematics in the Faculty of Natural Sciences at Warsaw University. With Łukasiewicz also a professor, Warsaw at this time had two more professors of mathematical logic than anywhere else in the world, and they founded the Warsaw School of logic. Leśniewski's students included many of Poland's finest logicians, most notably Alfred Tarski, who was the only student to complete a doctorate under him. Having at first rejected symbolic logic, from 1920 he embraced it and began to formulate his systems with unprecedented precision. Through the 1920s he developed and improved his system of foundations, and began publishing results towards the end of the decade in two important series of papers. In the 1930s he developed serious doctrinal disagreements with colleagues, most notably over his vehement opposition to set theory, and he became personally isolated as well as increasingly anti-Semitic in his views. He was diagnosed with thyroid cancer and after an unsuccessful operation he died on 13 May 1939 at the age of 52. His papers, including an unfinished book on logical antinomies, were destroyed in the 1944 Warsaw Uprising. Leśniewski's logic was unorthodox, uncompromising and inconvenient, so did not become mainstream, but his strictures on proper use of quotation, the distinction between object-language and

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

metalanguage, his canons of acceptable definition, and his principles of mereology (part-whole theory) passed into standard use. He was the first logician clearly to value Frege's work above that of Whitehead and Russell. Leśniewski left a small coterie of devoted followers who after World War II reconstructed as much as they could of his lost work and themselves went on to distinguished logical careers.

2 EARLY WRITINGS

In his early work Leśniewski wrote about aspects of the philosophy of logic, in particular the notion of truth and the status of the laws of logic. His work was influenced by Mill, Marty, Husserl and Cornelius, and was especially concerned with logical grammar. At this stage he worked with Mill's distinction between the objects a term denotes and the properties it connotes. In "A Contribution to the Analysis of Existential Propositions" (1911), based on his dissertation, he held that because all subject–predicate propositions can only be true if the subject term denotes, all positive existential propositions such as 'Lions exist' are analytically true, and all negative ones such as 'Unicorns do not exist' are contradictory. The logically proper way to formulate the true proposition that there are no unicorns is 'No being is a unicorn'. In "An Attempt at a Proof of the Ontological Principle of Contradiction" (1912) he argues that 'No object is able to be both B and not-B' is true but disagrees with the equivalence (now generally accepted) that 'No A is B' is equivalent to 'If something is A then it is not B'. In "The Critique of the Logical Principle of the Excluded Middle" (1913) he distinguishes between an ontological principle — *Every object is either A or not A* — and a logical principle — *Of two contradictory propositions, at least one must be true.* He rejects the latter because he takes 'Every centaur has a tail' and 'Some centaur does not have a tail' to be contradictories but both false because the subject-term 'centaur' is empty. In the course of this paper he argues against Twardowski's theory of general objects and Meinong's theory of impossible objects, as well as offering solutions to the paradoxes of Grelling and the Liar. This paper is the richest of his early "logico-grammatical" works. In "Is truth only eternal or both eternal and without a beginning?" (1913) he argues for timeless bivalence against Kotarbiński's contention that future contingent propositions lack a determinate truth-value. In "Is the class of classes not subordinate to themselves subordinate to itself?" (1914) he offers an analysis of Russell's Paradox, claiming that 'the class of As' refers to the unique mereological sum of As, so that since every object is subordinate to itself, no class of objects is not subordinate to itself, so Russell's Paradox fails to arise. During World War I, when Leśniewski lived in Moscow, he completed "Foundations of a General Theory of Sets" (1916). Despite the use of the term '*mnogość*' (set), this was a first rigorous deductive presentation of the theory of parts, wholes and concrete collections, for which, in order to differentiate his view from "official" set theory, Leśniewski coined the term 'mereology'. In this paper the language is a highly regimented Polish supplemented with variables, because at the time he did not trust symbolic methods.

3 FORMALIZATION

In 1920 Leon Chwistek persuaded Leśniewski to overcome the aversion to symbolic methods in logic that he had acquired principally through reading Whitehead and Russell's *Principia mathematica*, which was careless about the distinction between use and mention in a way which Leśniewski instinctively was not. From his very first paper Leśniewski made the distinction perfectly clearly, and he literally failed to understand works where confusion reigned. By contrast Leśniewski fully appreciated the care Frege took with the metalinguistic description of his logic.

In the course of the 1920s Leśniewski submitted his rigorous but unformalized reasonings regarding propositions, objects, collections and parts to a process of progressive formalization. The axioms of his 1916 mereology were rendered into a formalized axiomatic mereology, where undefined expressions meaning 'part of' and 'collection of' were taken as primitive. This rendering of mereology and its immediate consequences suffered two drawbacks: the expression for 'collection of' needed to be defined in terms of 'part of' before being inserted into the axioms; and the logical principles governing names and name-forming functors were not made explicit. The former was easily remedied by a new axiomatization; the latter required another logical system, one dealing with names, predicates, and other functors. This system, which picked up where the early papers left off, but now ditched the notions of connotation and property, Leśniewski came to call 'ontology', because he saw it as a general theory of objects in the sense of Aristotle. But ontology had its own logical presuppositions, a logic of propositions or "theory of deduction" as Whitehead and Russell had called it. This most basic theory of logical first principles Leśniewski called 'protothetic'. For him it included not just the standard truth-functional connectives, but also higher functors and the basic principles governing the logic of quantifiers.

Having established that the logical order of systems should be (1) protothetic (2) ontology (3) mereology, Leśniewski now set about giving each system an axiomatic formulation. For protothetic he was aided by a discovery of Tarski that conjunction is definable in terms of material equivalence and universal quantification, using propositional functions. This enable Leśniewski to base protothetic on just one connective and one quantifier. The advantage of having equivalence as primitive was that it allowed the simple formulation of definitions, which Leśniewski construed as object-language equivalences rather than metalinguistic abbreviations. For ontology he chose the singular 'is' (requiring no article in Polish) and found a single axiom for ontology too. For mereology the original primitive 'part of' was only one of many possible, and he experimented with different primitives.

Because Leśniewski's logical terms had their origins in his careful vernacular before he came to formalize them, he always conceived of his axioms and their constituent expressions as having definite meanings, as being true of the actual world, and the inference rules as truth-preserving. He steadfastly rejected all notions of formalization as involving the manipulation of symbols without meaning, and with deliberate irony called his approach "intuitionistic formalism", because

he was bent on rigorously formalizing reasoning with intuitively true axioms and with rules intuitively preserving truth.

4 NOMINALISTIC METALOGIC

This did not stop Leśniewski from exercising utmost rigour in formulating the principles of his system. But in this too he was distinctly unorthodox. Because he rejected platonism, Leśniewski conceived of his logical systems not as sets of abstract propositions or ideal sentence-types, but as concrete collections of physical complexes, growing in time by the addition of new theses. Since no such system is ever infinite, yet the ways in which they might grow, consonant with logical practice, are not given in advance, Leśniewski was forced to give directives which say when it is permitted to add a new thesis to a given system at any given stage. This meant that he in effect had to anticipate in advance and describe precisely but schematically the general form that any thesis could take in relation to prior theses. It was to the formulation of these metalogical directives that he gave his greatest effort.

In the course of formulating his directives, Leśniewski introduced a rich metalogical vocabulary, the terms of which were linked by a series of metalogical definitions he called 'Terminological Explanations'. The complexity of these was such that he was obliged to abbreviate them in the language of ontology, enriched with a number of primitive metalinguistic terms such as 'word', 'expression, 'first word of expression', 'parenthesis symmetric to', and more. The terminological explanations and directives for protothetic run to many pages, containing initial explanations, 49 terminological explanations and five directives for extending the system. The definition of 'definition of protothetic' alone runs to 18 clauses, and was the single achievement of which Leśniewski was proudest. It was the sloppiness of previous accounts of definition, with the exception of Frege, that prompted Leśniewski to formulate directives for appropriate definitions. These include that the expression defined occur only once on one side of an equivalence, with universally quantified variables, that the definiens employ only expressions previously defined, and so on. It sounds simple but in fact is not. In his advanced graduate seminars Leśniewski typically took three semesters to present the directives, proving by countermodels that every clause of every terminological explanation is logically independent of every other. No logician before or since took such trouble to be so scrupulously exact in their logic. The force of the directives is not imperative or prescriptive. There is no injunction to extend any logical system. Suppose a logical system (concrete collection of inscriptions) has been constructed according to the directives, starting from an inscription of the axiom and proceeding according to the directives up to a certain thesis. Now a logician takes it upon herself to extend the system by adding a new inscription. What the directives do is give the conditions under which any newly added inscription is a legitimate extension. They do not tell the logician how she should extend the system, which is a matter of will and ingenuity. Assuming the axiom is true, it can be seen

intuitively that any thesis added according to the directives will preserve truth if all the preceding theses are true. In other words, the system is sound. Leśniewski strove (successfully) to achieve this, but did not prove it.

5 SEMANTIC CATEGORIES

The system of directives and terminological explanations for his systems forced Leśniewski to think about the ways in which expressions can be formed out of others. In doing so he built on and modified Husserl's concept of meaning categories and Whitehead and Russell's theory of types. The result is that every expression (token, always) except punctuation (brackets, dots, commas etc.) has a particular semantic *category*. Categories are either basic or functorial. The basic categories in Leśniewski are SENTENCE (in protothetic) or SENTENCE and NAME (in ontology and mereology). All other categorematic expressions are *functors*, whose role is to combine with arguments of particular categories in sequence to determine a complex of a given category. Leśniewski's practice is modelled on that of Frege. It was later codified by Ajdukiewicz and formed the basis of categorial grammar. In Leśniewski's system there is only one quantifier, the universal, and it is syncategorematic, being expressed merely by putting a string of variables within special brackets (lower corners). He was unable to come up with general terminological explanations and directives for introducing new quantifiers, and despite his offer of any degree whatever to any student able to do so, no one, not even Tarski, succeeded.

While Leśniewski was punctilious about the form of his logical systems, the actual shape of the words was a matter of relative indifference provided no expression could be construed as compounded in two ways. He therefore laid down no specifications as to what shape expressions should have after the axioms, except for requiring symmetric brackets of different shapes to help determine categories. Unlike Church he did not require categorial indices for his variables.

The strictly regulated axiomatic systems specified by the directives were good for developing inscriptional metalogic but too unwieldy for normal use, and in everyday situations Leśniewski used a kind of natural deduction, similar to systems later developed by Jaśkowski and others, but never explicitly formulated by Leśniewski, who considered such deductions as mere sketches indicating how a full-dress axiomatic derivation should go. For metalogical purposes his formal language has all functors preceding their arguments enclosed in categorially suitable parentheses, and quantifier scope marked by upper corners, but again for everyday purposes he is happy to use infix notation, dots for parentheses, and a quantifier dual to the universal. Neither Leśniewski nor anyone else has formalized his implicit natural deduction rules nor shown in explicit detail how deductions performed according to them can be translated into axiomatic deductions according to the directives.

Leśniewski has peculiarly stringent desiderata concerning the axiology of axiom systems. A good axiom system should as far as possible have a single new primitive

term, a single axiom for it, which should be organic, i.e. not be a conjunction of parts which are themselves theorems, and finally it should be as short as possible. The pursuit of perfection according to these desiderata, which he shared with Łukasiewicz and which they passed on to their students, took up much energy among interbellum Polish logicians.

At some time in the early 1920s Leśniewski came to espouse extensionality. This is not a single principle but a form of principles. Two basic expressions are coextensional when they denote the same objects if they are names or are materially equivalent if they are sentences. Expressions of functor categories are coextensional when they always yield coextensional output values for coextensional input values, whatever these are. To guarantee that all expressions in the categorial hierarchy are extensional Leśniewski formulates a directive stating that once a new category of expression is introduced it is legitimate to add a principle of extensionality for that category. Expressions of new category are introduced by definitions: the axioms of a system typically contain only very few categories of expression.

6 MATURE WRITINGS

In the formative 1920s results, development, simplifications and other advances came so thick and fast, and with the collaboration of so many logicians, most notably Tarski, that Leśniewski never managed to put together a definitive account of his logical systems as a "systematic compendium", although he had planned to do so. To ensure that something was published and his results did not remain hearsay, he instead elected to publish papers setting out his results as an "autobiographical sketch", giving the basics as lucidly as possible, and indicating how he and others had chipped in with various results. The first result was the six-article Polish series 'O podstawach matematyki', 'On the foundations of mathematics', published 1927–31 in *Przegląd Filozoficzny*. The major part of this updates Leśniewski's mereology from the 1916 article, in a similar style, but with a wider range of offerings for possible primitives. It is prefaced by some autobiographical remarks indicating how Leśniewski came to mathematical logic, and adds a long, tedious and pedantic rant against the use/mention confusions in Whitehead and Russell, which Leśniewski says it took him four years to penetrate to give sense to their formulas. There is a discussion of Russell's Paradox and different ways of understanding the notion 'class', in which Leśniewski takes issue with Zermelo, Sierpiński, Hausdorff and other purveyors of "standard" set theory for embracing the empty set, distinguishing a singleton set from its member, and supposing that the only members (parts) of the set (mereological sum) of *A*s must be *A*s. A final Chapter on singular propositions of the type *A is (a) b* explains how Leśniewski arrived at his understanding of these propositions and came to the original axiom of ontology in 1920. It finishes with a discussion of tensed propositions in which Leśniewski argues that the proper subjects of propositions with tensed copulas are temporal parts of the whole things denoted by the unrestricted terms, so that

'Warsaw of 1930 is Warsaw of 1830' is false because the two terms 'Warsaw in 1930' and 'Warsaw in 1830' denote different objects, each of which is a different time-slice of the whole temporally extended object denoted by 'Warsaw'. It is one of Leśniewski's few excursions into metaphysical topics.

In 1929, Leśniewski published the first part of a long German article 'Grundzüge eines neuen Systems der Grundlagen der Mathematik', 'Fundamentals of a New System of the Foundations of Mathematics', dealing with protothetic. The first eleven sections appeared in *Fundamenta mathematicae*, then there was a hiatus of nine years, because Leśniewski withdrew the continuation after falling out with Wacław Śierpiński and the other mathematical editors of *Fundamenta* about his attitude to set theory. A new journal *Collecteana Logica* allowed him to publish the continuation in 1938, prefaced by a long sequence of introductory remarks updating the results of 1929 and attacking Johann von Neumann (not for the first time) for sloppy metalogic. The 'Grundzüge' offer the most complete published vision of how Leśniewski thought logic should be done. They contain an axiomatization in Whitehead–Russell notation account of the pure equivalential calculus of propositions, in which Leśniewski proves the metatheorem that a proposition of this calculus is a theorem if and only if every variable occurs an even number of times. Leśniewski sketches the history of simplifications and improvements to protothetic since its inception, then gives his Terminological Explanations and Directives for a system in his own notation. There the first article ended. In 1938 there followed a typesetter's nightmare: an unbroken sequence, scores of pages long, of 12 definitions and 422 theorems in Leśniewski's own unconventional notation, with only terse marginal indications of the prior theses employed. In the face of this uncompromising barrage there is little wonder that very few were prepared to make the effort needed to appreciate Leśniewski's work. Like Frege's before him, his notation has its merits, but it was documentation for the record, not an advertisement for adherents.

Apart from these long works Leśniewski published four shorter ones, also in German: an almost undecipherably dense set of Terminological Explanations for ontology, a more accessible investigation of the role of definitions in Łukasiewicz's propositional calculus, and two more dense results: single-axiom axiomatizations for the theories of groups and Abelian groups. These grew out of Leśniewski's teaching in Warsaw. Despite the loss of his posthumous manuscripts, some student notes survived the Second World War and a collection of edited English translations appeared as *S. Leśniewski's Lecture Notes in Logic* in 1988. Three sets deal with each of the three basic systems and contain material not covered in the published works, such as a sketch of an algorithmic calculus of truth-values in protothetic, a long list of theses and definitions in ontology, and a similar list for mereology. Other notes cover Peano arithmetic, inductive definitions, and a critical comparison of Whitehead's 1919 theory of events with mereology. The list (published in the 1988 volume) of seminars and courses Leśniewski held in Warsaw from 1919–39 show him ranging somewhat more widely, including courses on semantic antinomies and on axioms of geometry.

7 PROTOTHETIC

Prototheticis Leśniewski's propositional logic but it is much more than propositional calculus: it contains universal quantification and functor categories, initially of first level, but potentially up to any level desired. The primitive functor first chosen by Leśniewski is material equivalence, and he develops some of the pure equivalential calculus, based on the axioms of transitivity and associativity of equivalence

A1. $((p \leftrightarrow r) \leftrightarrow (q \leftrightarrow p)) \leftrightarrow (r \leftrightarrow q)$

A2. $(p \leftrightarrow (q \leftrightarrow r)) \leftrightarrow ((p \leftrightarrow q) \leftrightarrow r)$

together with rules of substitution and detachment. But from the first his interest is in the whole calculus. Negation is defined as

Def.$\sim$ $\forall p^{\ulcorner} \sim p \leftrightarrow (p \leftrightarrow \forall r^{\ulcorner} r^{\urcorner})^{\urcorner}$

while conjunction can be defined, following Tarski, as

Def.$\wedge$ $\forall pq^{\ulcorner} p \wedge q \leftrightarrow \forall f^{\ulcorner} p \leftrightarrow (f(p) \leftrightarrow f(q))^{\urcorner\urcorner}$

By taking the universally quantified closures of the equivalential axioms A1 and A2 and adding a third axiom

A3. $\forall gp^{\ulcorner} \forall f^{\ulcorner} g(pp) \leftrightarrow (\forall r^{\ulcorner} f(rr) \leftrightarrow g(pp)^{\urcorner} \leftrightarrow \forall r^{\ulcorner} (f(rr) \leftrightarrow g((p \leftrightarrow \forall^{\ulcorner} q^{\urcorner}) \leftrightarrow p)^{\urcorner}) \leftrightarrow \forall q^{\ulcorner} g(qp)^{\urcorner\urcorner\urcorner}$

Leśniewski obtained an early basis for prototheticLike all his logics, prototheticwas subject to successive simplifications, as hardly anyone could be expected to understand the import of such a formula, still less the axioms of single-axiom systems into which Leśniewski and his students poured effort and ingenuity. Multiple-axiom systems are more perspicuous. Here is one using the material conditional as the basic connective

P1. $\forall pq^{\ulcorner} p \rightarrow (q \rightarrow p)^{\urcorner}$

P2. $\forall pqr f^{\ulcorner} f(rp) \rightarrow (f(r \forall s^{\ulcorner} s^{\urcorner}) \rightarrow f(rq))^{\urcorner}$

The directives obviously have to be reformulated for the conditional as basis.

In its logical force, prototheticis equivalent in potential to a theory of propositional types, such as described much later by Leon Henkin. Because of the extensionality of the system, there are only finitely many possible semantic values for the constants of each type or semantic category, and every universal quantification is logically equivalent to a conjunction of formulas instantiating all the possible combinations of values for the variables quantified. Because of this, the chief role of the quantifiers is abbreviatory: they shorten formulas and obviate the need to formulate new directives for each new category of functor introduced by definition. Their introduction paves the way for their use in later systems, where they are not finitely eliminable.

8 ONTOLOGY

Ontology is the second of Leśniewski's logical systems and of greater philosophical and logical interest. Leśniewski describes it as a kind of amalgam of traditional logic with Schröder's calculus of classes with individuals, informed throughout however by the rigour of Frege. Ontology adds a new primitive category of NAME to the SENTENCE of protothetic, and the rules and directives for quantifiers, definitions and substitution are extended to cope with the new swathe of categories made possible. The names of ontology need not be singular, but can be empty or plural. This has implications for the interpretation of quantifiers: these cannot be understood as standardly in predicate logic, about which more below.

Leśniewski elected to formulate ontology in terms of the functor 'ε' of singular inclusion. He followed Peano in selecting the first letter of the Greek word '$\varepsilon\sigma\tau\iota$', 'is', and typically rendered sentences of the form '$a\,\varepsilon\,b$' in Polish as 'a jest b'. Since Polish lacks articles, there is no unique best way in which to say this in English: taking the options into account, one might cover several by '(the) a is (a) b', but for logical purposes a more honest rendering might be '(the) a is one of (the) bs', which makes it clear that it is inclusion of one object among one or more. The truth-conditions of a formula '$a\,\varepsilon\,b$' are straightforward: it is true if 'a' denotes a single object and this object is one of the one or more objects denoted by 'b'; in all other cases the formula is false. But while the simplest rendering into English is 'a is a b', it should not be thought that '$\,\varepsilon\,$' is a mere copula. The mistake is sometimes made of attributing to Leśniewski a two-name account of predication, like that of Ockham. This is wrong: Leśniewski understood predication much as Frege did, involving an unsaturated predicate and one or more names. 'ε' is not a mere copula, but a two-place predicate out of the same syntactic and semantic category as other relational predicates like 'admires': 'John admires mountaineers' and 'John admires the Beatles' are grammatically exactly like 'John is a mountaineer' and 'John is one of the Beatles'. A one-place predication like 'John smokes' or 'Lions roar' both have the same form for Leśniewski: $f(a)$. The major difference is that Leśniewski, following the tradition from Aristotle to Schröder, and unlike post-Fregean orthodoxy, puts general terms in the same category as singular terms.

In view of the meaning of '$a\,\varepsilon\,b$' Leśniewski sought to spell out its logical force in an axiom which he modelled on Russell's analysis of definite descriptions: (the) a is a b if and only if there is at least one a, there is at most one a, and every a is a b. Putting the parts of this equivalence all in terms of 'ε' yields the following axiom

AO $\quad \forall ab\ulcorner a\,\varepsilon\,b \leftrightarrow (\exists c\ulcorner c\,\varepsilon\,a\urcorner \wedge \forall de\ulcorner (d\,\varepsilon\,a \wedge e\,\varepsilon\,a) \rightarrow d\,\varepsilon\,e\urcorner \wedge \forall f\ulcorner f\,\varepsilon\,a \rightarrow f\,\varepsilon\,b\urcorner)\urcorner.$

The form of this, with the undefined term occurring once on the left-hand side in an atomic sentence, and used on the right-hand side, amounts to a kind of implicit definition of 'ε'. Notice how the term 'a' occurs in subject position on the left but only in the complement position on the right. Subsequent investigations allowed this 1920 axiom to be replaced by shorter ones, such as the pair

AS1 $\forall ab\ulcorner a\,\varepsilon\, b \rightarrow a\,\varepsilon\, a\urcorner$

AS2 $\forall abc\ulcorner(a\,\varepsilon\, b \wedge b\,\varepsilon\, c) \rightarrow a\,\varepsilon\, c\urcorner$

but these short axioms are less obviously explicative of the meaning of 'ε' than the original long axiom.

In terms of singular inclusion a wide range of other logical predicates can be defined, involving existence and non-existence, inclusion, exclusion and overlapping, singular identity (which is mutual singular inclusion), other equivalences, including coextensiveness, and numerical predicates like 'there are exactly seven'. These are typically either generalized quantifiers or higher-order predicates in ordinary predicate logic, so that ontology already carries useful expressive power at the first-order or elementary level, where only nominal variables are quantified. In fact it is equivalent in logical strength to monadic second-order predicate logic, which is complete and decidable.

In addition to predicates and other propositive functors, ontology allows the definition of names and other nominative functors, for which a special directive of ontological definitions is required. To avoid antinomies like Russell's, both sides of an ontological definition have to be existence-implying. A case in point is the ontological definition of the standard empty name 'Λ':

Def.Λ $\forall a\ulcorner a\,\varepsilon\,\Lambda \leftrightarrow (a\,\varepsilon\, a \wedge \sim (a\,\varepsilon\, a))\urcorner$

Leaving out the first conjunct in the definiens would give a definition of a Russell-like antinomic name, but the existence-implying first conjunct ensures the name defined is consistent, merely necessarily empty.

Because of its traditional use of nominal terms, ontology makes a more elegant and natural vehicle than predicate logic for the modern representation of reasoning in traditional syllogistic. It also fits more neatly with much of 19th century logic, from Boole and De Morgan through Jevons, Brentano and Lewis Carroll to Schröder.

Beyond elementary ontology, when predicate and other functors variables are quantified, it becomes possible to define mathematically interesting notions such as one–one correspondence or cardinal equality, cardinal less-than, finitude and infinitude, and cardinal number, as well as formulate principles of choice and replacement, and continuum hypotheses. With some exceptions among the Notre Dame school of Leśniewski's former pupil Sobociński however, such excursions into higher types, or higher semantic categories, have been regrettably rare. Because of the indefinitely extensible nature of ontology, its potential matches that of a simple theory of types, but unlike Whitehead and Russell, Leśniewski did not introduce schematic notation allowing the expression of type- or category-neutral thesis schemata. Little of Leśniewski's published work concerns ontology, which is a pity as it is in many ways his most original and interesting creation. Some idea of what he could do with it can be gleaned from the packed but terse "Definitions and Theses of Leśniewski's Ontology" in the 1988 *Lecture Notes*.

9 MEREOLOGY

Of all Leśniewski's systems, it is mereology that is best known outside esoteric circles, if only because he invented the name. Its genesis as an antinomy-free nominalistic alternative to Cantorian set theory not only left it with class-like terminology, but also affected the part–whole principles Leśniewski embraced, which are motivated mathematically rather than metaphysically. The mereology is extensional: things are identical if they have the same parts; and compositionally universal: any (non-empty) collection of things composes another thing, their mereological sum or fusion, which by extensionality is then unique. There is thus a unique maximal thing, the sum of all things. The system is thus as generous with individual things as it is possible to be, compatible with extensionality and the lack of type-difference between things and sums.

More than in any other of his systems, Leśniewski rung the changes in the axioms and primitives of his mereology. The 1927-31 series 'O podstawach' in particular runs through many alternatives, and looks for single axioms. But again perspicuity is a virtue often incompatible with parsimony of primitives and axioms. The following four axioms give an idea of how mereology can be formulated perspicuously. Rather than use a relational predicate of part–whole, Leśniewski uses a name-forming functor of one nominal argument, here written 'pt[ ]' and naturally read as 'part of'. It includes the limiting case of improper part, whereby an individual is an improper part of itself. The second functor 'Sm[ ]' from the same category means '(the) Sum of', the capital letter indicating uniqueness:

M1 $\quad \forall ab^{\ulcorner} a \ \varepsilon \ \mathrm{pt}[b] \rightarrow b \ \varepsilon \ b^{\urcorner}$

M2 $\quad \forall abc^{\ulcorner}(a \ \varepsilon \ \mathrm{pt}[b] \wedge b \ \varepsilon \ \mathrm{pt}[c]) \rightarrow a \ \varepsilon \ \mathrm{pt}[c]^{\urcorner}$

M3 $\quad \forall ab^{\ulcorner} a \ \varepsilon \ \mathrm{Sm}[b] \leftrightarrow (a \ \varepsilon \ a \wedge \forall c^{\ulcorner} c \ \varepsilon \ b \rightarrow c \ \varepsilon \ \mathrm{pt}[a]^{\urcorner} \wedge$
$\forall d^{\ulcorner} d \ \varepsilon \ \mathrm{pt}[a] \rightarrow \exists e f^{\ulcorner} e \ \varepsilon \ b \wedge f \ \varepsilon \ \mathrm{pt}[e] \wedge f \ \varepsilon \ \mathrm{pt}[d]^{\urcorner\urcorner})^{\urcorner}$

M4 $\quad \forall ab^{\ulcorner} a \ \varepsilon \ b \rightarrow \mathrm{Sm}[b] \ \varepsilon \ \mathrm{Sm}[b]^{\urcorner}$

Here M1 says that if a is a part of b then b is an individual, M2 says that parthood is transitive, M3 defines the unique Sum of the bs as the smallest thing which overlaps every b, and M4 says that if there is a b then the Sum of the bs exists and is an individual.

The mereological sum of the bs differs from the class of the bs in several ways. If there is only one b, the Sum *is* that b. If the bs are concrete (which for the nominalist Leśniewski they always are) then so is the Sum. And not every part of the Sum of the bs is necessarily a b. If the bs are books, for example, one part of the Sum of all books is the Sum of all book spines, but this is no book. There is no Sum of nothing, or null Sum. Finally, the Sum of the Sum of all bs is just the Sum of bs again, since this is an individual, and the Sum of an individual is itself. So no hierarchy of Sums can get started, which is very unlike the case for sets.

For these reasons mereology is very different from standard set theory. It has finite models, for instance the model of a single atom (individual with no proper part) object, and so is consistent. All its finite models have cardinalities of the form $2^n - 1$, where n is the number of atoms in the model. Mathematically, mereology is like Boolean algebra extended with infinitary meet and join operators, except that it lacks a zero or null element. Despite Leśniewski's original intention to use mereology as part of a foundation for mathematics, it is far too weak a theory to do as much work as for example Zermelo-Fraenkel set theory.

It is independent of general mereology whether or not there are atoms, and if there are atoms, whether everything is composed of them or not. There are three possible extensions of general mereology to account for these three possibilities. Only atomistic mereology has finite models, but there may be infinitely many atoms even in an atomistic mereology. Atomless mereology models a world of "gunk" in which everything has a proper part.

10 PHILOSOPHICAL ASPECTS OF LEŚNIEWSKI'S SYSTEMS

Leśniewski trained as a philosopher and many of the more awkward and interesting aspects of his logic can be traced to his philosophical views. Leśniewski's nominalism has been noted. It is a controversial position, but not one on which he has anything very new to say. His argument against Twardowski's theory of general objects is no threat to a realist about universals. The general b is defined as that object which has all and only those properties had by all bs. Suppose there are some bs which are cs and some which are not, as will always be the case for some c if there is more than one b. The general b cannot have the property of being a c since not all bs are cs, and it cannot have the property of not being a c since not all bs are not cs. Hence by *reductio* there cannot be more than one b. A realists about universals will simply reject the definition. Leśniewski's colleague and friend Kotarbiński employed Leśniewski's logic in the service of his materialistic reism, the view that all that exists are individual material things. Leśniewski was sympathetic but did not publicly subscribe to or support Kotarbiński, because he could not work out how to fit non-material phenomena like after-images into the reistic picture.

Like nearly all students of Twardowski, Leśniewski followed the master in supporting a view of truth and falsity as absolute. He refused for example to relativize truth to time. It is unclear whether he rejected the very idea of a definition of truth, but he certainly rejected Tarski's famous definition, because of its use of the hated set theory. Leśniewski's vituperative objections to set theory aroused opposition in his own lifetime: they are based in part on his nominalism and in part on his occasionally idiosyncratic reading of Cantor and other set theorists. He was seemingly unable to come to terms with set theory as an axiomatic theory like any other in mathematics, and preferred to get by without it rather than sell out, as he no doubt thought Tarski had, for the sake of metamathematical results.

Leśniewski's understanding of quantifiers has come in for much commentary, because it is clearly different from the standard "domain-and-values" account practised by Frege and Russell, formalized by Tarski and epitomized by Quine. In terms of the deductive *role* of the quantifiers — while Leśniewski officially only has the universal quantifier '$\forall$', in everyday deduction he also uses the other, the *particular* quantifier '$\exists$' — there is nothing untoward, no restrictions on instantiation or generalization, in terms of *meaning* the variables cannot be understood to be ranging over a domain of entities. The clearest proof of this is in ontology. Through defining the standard empty name 'Λ' we are entitled to infer the following theorem

$$\exists a^{\ulcorner} \sim \exists b^{\ulcorner} b \, \varepsilon \, a^{\urcorner\urcorner}$$

which, if '$\exists$' meant 'there exist' would have to mean 'there exists something that doesn't exist', which would be absurd. It is not even right to read it as "Something does not exist", which would be the Meinongian view Leśniewski rejects. Rather we have to simply say "For some a, there exists no a". This is not just true, but a logical theorem, so cannot be absurd. The question is how to interpret the quantifier.

Further indication that Leśniewski cannot understand the quantifiers standardly comes from his attitude to ontology. A domain-and-values interpretation would have Leśniewski committed to truth-values and functions to them in protothetic, to properties, relations and functions of any order in ontology. But when Quine visited Warsaw in 1933 and tackled Leśniewski about the ontology to which his logic was committed, Leśniewski denied any such commitment. Indeed the very idea of ontological commitment may have originated in these discussions. Quine came away from them convinced that Leśniewski understood quantifiers substitutionally, so that quantifiers ranged over expressions rather than what they denote. But this cannot be right either. For the nominalist Leśniewski all expressions are concrete individuals, and there are fewer of them than there things in general. Of course it is expressions that are substituted for variables when quantified sentences are instantiated, but to replace or be replaced by a variable is not to be quantified over by it. Leśniewski indicated in his early writings that he understood a quantified expression such as '$\exists a.f(a)$' as saying "for some meaning (signification) of the expression 'a', $f(a)$", and attributes this reading, quite erroneously, to Peirce, who in the context in question would have said rather "f is true of some one of the individuals denoted by 'a"', which is just the orthodox understanding, with a minor detour via a mention of the variable. The fact is that for Leśniewski quantification is *sui generis* and not to be explained either via substitution or by domain and values. His own reading can be rendered adverbially as "Somehow that 'a' signifies, $f(a)$". We have to take it that expressions may signify alike or differently: sentences by being true or false, names by denoting these or those individuals, and functors by combining the ways their inputs and outputs signify. As an opponent of a semantics tacked onto a non-signifying language from the outside, Leśniewski could not accept that the quantifiers acquire a meaning only from a domain and interpretation, but there is still a gap in his explanation. Tarski's

use of infinite sequences to interpret quantifiers did not meet with his approval, because it employed abstract sets.

On another aspect of the philosophy of logic, Leśniewski is easier to understand. This is the view that logic should be ontologically neutral, that is, that no logical thesis should commit us to the existence of anything. In other words, the theses should be true even were there to be no individuals. This means all theses should be valid for the empty universe, a condition not satisfied by standard predicate logic. Of course for a nominalist the empty universe is empty of anything, including higher-order or abstract entities. For a platonist the empty universe is devoid merely of concrete individuals.

Leśniewski's rabid extensionalism — in which he did not differ from his Polish contemporaries — meant that he could not accept intensional contexts or a Fregean distinction between sense and reference. In the published works he says very little about this: apparently in the 1930s, according to a 1935 letter to Twardowski, he worked on intensional operators as well as paradoxes, but such work as he did was incomplete and was lost with the rest of his *Nachlass* in 1944. The other lost work of the 1930s is an examination and critique of Łukasiewicz's systems of many-valued logic.

11 LEŚNIEWSKI'S INFLUENCE AND SIGNIFICANCE

For many reasons — Leśniewski's early death, the loss of his manuscripts, the devastation of his country, and the deaths and diaspora of Polish logicians in World War II — the influence of Leśniewski has been more muted than his originality and brilliance would lead one to expect. In Poland, Jerzy Słupecki publicised and continued his work, while abroad former pupils such as Sobociński and Lejewski did likewise. But they were isolated and out of step with the times. For this Leśniewski himself is partly responsible. Some of the reasons were astutely summarized by Grzegorczyk in 1955. Up until the late 1920s, the investigation and improvement of logical systems "from the inside" was the way logic progressed, but as the threat of antinomies gradually receded, attention shifted towards the metamathematical investigation of systems and to model theory, in which Tarski played a major part. Leśniewski's conception of logical systems as growing in time and as always unfinished made them, as Tarski put it in a footnote to his 1933 paper on truth, "thoroughly unsuitable objects for metamathematical research" (a passage omitted from the English translation of Tarski's paper). Leśniewski however, through his hiatus in publishing and his intransigent opposition to set theory as a tool of investigation, placed himself outside and indeed against these developments. His influence with mathematicians was therefore nil. The limitative theorems of Gödel, which Leśniewski initially found hard to believe, put paid to the ideal of a once-and-for-all "true foundational system" that Leśniewski had been striving for. Axiomatic methods were replaced in elementary logic and its pedagogy after 1945 by natural deduction, so Leśniewski's efforts in axiomatics appeared not just difficult and old-fashioned but to some extent pointless. The

utter rigour and obsession with detail that Leśniewski brought to the formulation of systems came to seem idiosyncratic and myopic by comparison with the larger developments in logic and mathematics. The turning of the tide can be seen clearly in Tarski's famous long truth paper, whose early sections on antinomies and quotation, object-language and metalanguage acknowledge indebtedness to Leśniewski; but while Tarski initially stayed within the bounds of finite types, he then dissociated himself from the limitation and embraced a type-free approach which gained wider recognition. Add to this Leśniewski's unfortunate knack for alienating people and one gets the picture of a brilliant but grumpy eccentric left behind by trends he was unwilling to follow.

There is some truth in this picture, but it is only partial. Logicians have always had difficulty gaining recognition among mathematicians: even Tarski was less successful in "selling" logic to mathematicians than he would have liked, and Frege's deeply unsuccessful career was in many ways much more startlingly at variance with his genius than Leśniewski's. Leśniewski's life's work was incomplete, but it did not end in failure like Frege's logicism, and unlike Frege he had good students and saw several progress in their careers, in particular Tarski, of whom Leśniewski boasted "For doctoral students I have 100% geniuses." Leśniewski's role in establishing Warsaw as the world centre for logic was acknowledged in his day and to his face, and his achievements in the theory of definition, in mereology, in categorial syntax and in anchoring the object-language/metalanguage and use/mention distinctions are irreversible advances. While mereology is now absorbed into the ontological mainstream, and protothetic is a mere historical curiosity, his most interesting system, ontology, still remains well short of realizing its potential.

What value one places on Leśniewski's more idiosyncratic and less widely adopted positions depends on one's philosophical standpoint. For a platonist, Leśniewski's obsessions are merely whimsical, but for a nominalist they are a devastatingly honest confrontation of the rigours of eschewing abstract comforts. Leśniewski was not a mathematical genius like Tarski, nor was he an urbane leader and stylist like Łukasiewicz or a literate public diplomatist like Kotarbiński. He was indeed an obsessive, with ultra-high standards of rigour, precisely the sort of obsessive one would wish to have writing computer programs controlling nuclear power stations, running space missions or encrypting internet payments. It is no accident that Leśniewski was an adept *déchiffreur* of Soviet military code. Quine — tellingly, a sincere admirer — wrote his epitaph when he said "Lesniewski was notable for the degree of prolixity which he was willing to admit in the interest of complete rigor and precision." He was at the heart of the Warsaw School: without him it would have been very different and much less considerable. And as long as there are nominalists, the final word on Leśniewski's influence has yet to be spoken.

BIBLIOGRAPHY

Primary Literature

Leśniewski published his work in Polish, Russian, and German, and it has mainly not been republished. The only collected edition of his work is in English translation:

[Leśniewski, 1992] S. Leśniewski, *Collected Works*. In 2 volumes. Edited by Stanisław J. Surma, Jan T. Szrednicki and D. I. Barnett, with an annotated bibliography by V. Frederick Rickey. Dordrecht: Kluwer, 1992.
This collection contains all of Leśniewski's published works except book reviews. It does contain a fair number of typographical errors. The bibliography lists the original sources from which the translations are taken, as well as comprehensive secondary literature, but runs only to 1978.

[Leśniewski, 1988] *Stanisław Leśniewski's Lecture Notes in Logic*. Edited by Jan T. J. Szrednicki and Zbigniew Stachniak. Dordrecht: Kluwer, 1988.
These were reconstructed from notes taken by Leśniewski's students.

Secondary Literature

Secondary literature is scattered across journals, but the most useful sources are two collections:

[Srzednicki and Stachniak, 1988] J. T. J. Srzednicki and Z. Stachniak, eds. *Leśniewski's Systems: Protothetic*. 92. Dordrecht: Kluwer, 1998.
Contains essays by Rickey, Simons, Słupecki, Sobociński, and Tarski.

[Srzednicki and Rickey, 1984] J. T. Srzednicki and V. F. Rickey, eds., assistant editor: Czelakowski. *Leśniewski's Systems: Ontology and Mereology*. The Hague: Nijhoff, 1984.
Contains essays by Clay, Iwanuś, Kruszewski, Lejewski, Słupecki, and Sobociński.

[Luschei, 1962] E. C. Luschei, *The Logical Systems of Lesniewski*. North-Holland, 1962.
A detailed and militantly sympathetic exposition of Leśniewski's views, with emphasis on his nominalism and the terminological explanations.

[Woleński, 1989] J. Woleński. *Logic and Philosophy in the Lvov-Warsaw School*. Dordrecht: Kluwer, 1989.
This is an indispensable source for Leśniewski's position in the context of Polish logic and philosophy.

Internet resource

The Polish Philosophy Page at `http://www.fmag.unict.it/~polphil/polphil/MainPhil.html` contains biographical and bibliographical background to Leśniewski and his contemporaries, as well as revealing and otherwise unpublished correspondence with Twardowski and excerpts from others' remarks on Leśniewski.

HILBERT'S PROOF THEORY*

Wilfried Sieg

> Überhaupt ist es ein wesentlicher Vorzug der Mathematik gegenüber der Philosophie, dass sie sich mit ihrer Spekulation niemals in phantasievolles mystisches Denken hüllt.[1]

1 CHANGING PERSPECTIVES

Hilbert gave lectures on the foundations of mathematics throughout his career. Notes for many of them have been preserved and are remarkable treasures of information; they allow us to reconstruct the circuitous path from Hilbert's logicist position, deeply influenced by Dedekind and presented in lectures starting around 1890, to the novel program of finitist proof theory in the 1920s. The general direction of this path in quite unchartered territory was determined by Hilbert's focus on the *axiomatic method* and the associated *consistency problem*. Thus, our journey along that path begins with an analysis of the modern axiomatic method, proceeds to a discussion of the role of consistency proofs, and ends with a sketch of the content of Hilbert and Bernays' magisterial work *Grundlagen der Mathematik*. What a remarkable path it is: emerging as it does from the radical transformation of mathematics in the second half of the 19^{th} century, it leads to the new field of metamathematics or mathematical logic, with Hilbert insisting from the start that

*This essay is dedicated to George Boolos. I got to know George in 1978 and treasured from the very beginning his openness, support, and warmth. We shared logical interests (proof theory, albeit working in quite different parts) and historical/philosophical ones (emergence of modern logic and related investigations in the foundations of mathematics, albeit focusing on different aspects and people). The latter shared interest reveals an undercurrent of greater commonality: we were both fascinated by provably defunct programs, namely, Frege's logicist and Hilbert's finitist one, and asked "What is still of contemporary systematic interest?"

An initial version of this essay was prepared for the Boolos Conference at Notre Dame and delivered on April 16, 1998 under the title *Hilbert's Consistency Proofs*. Parts of the talk were expanded and published separately, for example, in *Sieg 2002*. On the other hand, parts of this essay are based on work I have done since. In particular, section 2 is informed by work I did for my seminar *Structuralism and its historical roots*, given at Carnegie Mellon in the spring of 2001, whereas sections 4 and 5 make use of parts 3, 4 and 5 of *Sieg & Ravaglia*.

My discussions with Mark Ravaglia during the years when he was working on his thesis, *Ravaglia 2003*, were very fruitful; his mathematical and philosophical analysis of the first volume of *Grundlagen der Mathematik* is still the most thorough and nuanced.

[1]Hilbert, *Mengenlehre*, Summer 1917, p. 102. The English translation is: In fact, it is an essential advantage of mathematics when compared to philosophy that it [mathematics] never wraps itself with its speculation in fanciful mystical thinking.

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

foundational problems be treated mathematically as far as possible. Examining that path with Hilbert and later on also with Bernays allows us to appreciate Hilbert's changing perspectives on the wide-open mathematical landscape, but it also enriches our perspective on Hilbert's metamathematical work. Though this essay sheds light on contemporary issues, it is intended to be mostly historical. Thus, I will quote extensively in particular from the unpublished source material.

1.1 A pivotal essay

In some sense, the development towards *Beweistheorie* really begins in late 1917 when Hilbert gave a talk in Zürich, entitled *Axiomatisches Denken*. This talk is rooted in the past and points to the future. Hilbert suggested in particular,

> ... we must — that is my conviction — take the concept of the specifically mathematical proof as an object of investigation, just as the astronomer has to consider the movement of his position, the physicist must study the theory of his apparatus, and the philosopher criticizes reason itself.[2]

Hilbert recognized full well, in the very next sentence, that "the execution of this program is at present, to be sure, still an unsolved problem." And yet, initial steps had been taken at the International Congress of Mathematicians in Heidelberg (1904); Hilbert presented there a syntactic consistency proof for a purely equational fragment of number theory. His 1905-paper, on which the talk was based, insisted programmatically on a simultaneous development of logic and number theory. Hilbert expected that his proof could be extended to establish the consistency for such a joint formalism.

Already four years earlier at the previous International Congress of Mathematicians in Paris, Hilbert had articulated the need for a consistency proof as his Second Problem. However, at that point the consistency of analysis was at stake, and it was to be established by semantic techniques adopted from the theory of real numbers. Neither in his Paris nor in his Heidelberg talk did Hilbert specify a suitable logical calculus; that also holds for all the lectures he gave before the

[2] *Ewald*, p. 1115. Here is the German text from *Hilbert 1918*, p. 155: ... [wir] müssen — das ist meine Überzeugung — den Begriff des spezifisch mathematischen Beweises selbst zum Gegenstand einer Untersuchung machen, gerade wie ja auch der Astronom die Bewegung seines Standortes berücksichtigen, der Physiker sich um die Theorie seines Apparates kümmern muß und der Philosoph die Vernunft selbst kritisiert.

This programmatic perspective is reiterated in Hilbert's *1922*, pp. 169-170: Um unser Ziel zu erreichen, müssen wir die Beweise als solche zum Gegenstande unserer Untersuchungen machen; wir werden so zu einer Art *Beweistheorie* gedrängt, die von dem Operieren mit den Beweisen selbst handelt. Für die konkret-anschauliche Zahlentheorie, die wir zuerst betrieben, waren die Zahlen das Gegenständliche und Aufweisbare, und die Beweise der Sätze über Zahlen fielen schon in das gedankliche Gebiet. Bei unserer jetzigen Untersuchung ist der Beweis selbst etwas Konkretes und Aufweisbares; die inhaltlichen Überlegungen erfolgen erst an dem Beweise. Wie der Physiker seinen Apparat, der Astronom seinen Standort untersucht, wie der Philosoph Vernunftkritik übt, so hat meiner Meinung nach der Mathematiker seine Sätze erst durch eine Beweiskritik sicherzustellen, und dazu bedarf er dieser Beweistheorie.

winter term 1917-18. It is only then that the study of *Principia Mathematica* bore fruit; the process of absorbing the core of Whitehead and Russell's work had begun in 1913. It provided a way of formulating joint formalisms with appropriate logicist and inferential principles, thus transcending the restricted systems of the Heidelberg talk. Hilbert had found the means to develop number theory and analysis, strictly formally.

Proof theory, as we understand it, begins only in 1922 when the finitist standpoint is providing the methodological grounding for consistency proofs, but the developments between 1917 and 1922 are extremely important. First of all, what was only alluded to at the end of the Zürich talk was actually carried out in Hilbert's remarkable 1917-18 lectures *Prinzipien der Mathematik*. Secondly, in his lectures from the summer term of 1920, Hilbert goes back to the syntactic perspective of the Heidelberg talk and reproves the consistency result I mentioned — in a slightly different way. Last but certainly not least, Paul Bernays joined Hilbert in the fall of 1917 as assistant for work on the foundations of mathematics. He not only helped prepare Hilbert's lectures on foundational issues and wrote the notes for many of them, starting with those of the winter term 1917-18, but he became a crucial collaborator in a joint enterprise and was seen by Hilbert in that light. It is difficult to imagine the evolution of proof theory without his mathematical acumen and philosophical sophistication. Indeed, my systematic presentation of Hilbert's proof theory culminates in a discussion of the second volume of *Grundlagen der Mathematik*, a volume that was written exclusively by Bernays.

The volumes of *Hilbert & Bernays* are milestones in the development of mathematical logic, not just of proof theory. They were at the forefront of contemporaneous research and presented crucial metamathematical results that are still central: from consistency proofs Hilbert and Bernays had obtained in weaker forms during the 1920s, through theorems of Herbrand and Gödel, to a sketch of Gentzen's consistency proof for number theory. The core material is expanded in the second volume through a series of important supplements concerning particular topics. One finds there, for example, an elegant development of analysis and an incisive presentation of the undecidability of the decision problem for first-order logic. In short, the two volumes constitute an encyclopedic synthesis of metamathematical work from the preceding two decades. What is most remarkable, however, is the sheer intellectual force that structures the books: these are penetrating studies concerned with and deeply connected to the foundations of modern mathematics as it emerged in the second half of the 19^{th} century.

1.2 Transformations

Do we have to go back that far in order to obtain a proper perspective on Hilbert's proof theory? Don't we have an adequate grasp of its development in the 1920s? Is there more to be said about the connection to the earlier foundational work than what was indicated above, supplementing the standard view of the evolution of proof theory? According to that common view, endorsed even by Bernays in

1935[3] and Weyl in *1944*, the evolution is divided into two periods. During the first period, roughly from 1899 to 1904, Hilbert is seen as unsuccessfully addressing difficulties for the foundations of arithmetic that had been raised by the set theoretic paradoxes. The second period begins around 1920, motivated again externally by the emergence of Brouwer's and Weyl's intuitionism, and it is seen as lasting either until 1931 (the publication of Gödel's incompleteness paper) or until 1939 (the publication of the second volume of *Grundlagen der Mathematik*). The initial work of the second period is characterized as a refinement of the earlier approach and as leading to Hilbert's *finitist consistency program*. Of course, it is recognized that this refinement involves a significant methodological shift and requires novel metamathematical investigations.

For both periods Hilbert's proposals are, however, *more* than defensive ad-hoc reactions against threats to "classical" mathematics and emerge out of long periods of intense work on difficult issues. The issues are connected in both cases to the creative freedom in mathematics celebrated by Cantor, Dedekind, and Hilbert himself, but seen as severely restricted by constructivity requirements that were articulated first by Kronecker and then by Brouwer. Thus, they are ultimately grounded in the 19^{th} century transformation of classical into modern abstract mathematics. Hilbert's problems and his metamathematical approaches are deeply intertwined with this transformation, and we can better understand it by better understanding Hilbert's perspective on it.[4]

The need for historical study and conceptual analysis is thereby not exhausted, as the precise character of Hilbert's problems and his techniques for solving them developed substantially and often dramatically. The complex evolution is chronicled in the notes for lectures on mathematics and logic that were given by Hilbert between 1890 and 1930. Toepell and Peckhaus have made clear how important these notes are for the early part of this long period. Hallett and Majer have strengthened that perspective through the first volume of the edition of *David Hilbert's Lectures on the Foundations of Mathematic and Physics, 1891-1933*; that volume presents and analyzes lectures on the foundations of geometry given between 1891 and 1902. My paper *Hilbert's Programs: 1917-1922* investigates the progression from a Russellian logicist program articulated in 1917 to the finitist

[3]In *Bernays 1935* there is a general smoothing of the developments between 1904/05 and 1917/22: Bernays does not mention that Hilbert gave lectures on the foundations of mathematics during that period; the 1917-18 lectures are not even hinted at. Bernays effectively creates the impression that the period is one of inactivity; e.g. on p. 200 one finds: In diesem vorläufigen Stadium hat Hilbert seine Untersuchungen über die Grundlagen der Arithmetik für lange Zeit unterbrochen. Ihre Wiederaufnahme finden wir angekündigt in dem 1917 gehaltenen Vortrage *Axiomatisches Denken*. The impression is reinforced by a footnote attached to the first sentence in the above quote, where Bernays points to the work of others who pursued the research direction stimulated by Hilbert's *1905*.

[4]Bernays emphasizes in his *1930* (on pp. 17-18 and pp. 19-20) that the changes in the methodological perspective of mathematics during the second half of the 19^{th} century had three characteristic features, namely, the use of the concept of set for the rigorous foundation of analysis, the emergence of existential axiomatics ("existentiale Axiomatik"), and building a closer connection between mathematics and logic. Existential axiomatics, the central concept, will be discussed extensively in section 2.

program of 1922: such a study would have been impossible without access to the lecture notes from those years. The historical evolution shows how deeply connected Hilbert's investigations are. Many fascinating aspects have been detailed or re-examined, for example, by Avigad, Corry, Ewald, Feferman, Ferreiros, Gray, Hallett, Kahle, Majer, Mancosu, Moore, Parsons, Ravaglia, Rowe, Tait, Tapp, and Zach.

1.3 Immediate context & overview

Hilbert's synthesizing perspective and the reasons for his insistence on formulating sharp problems when confronting the foundations of mathematics come to life in the long preamble to his Paris address. His account of mathematics emphasizes the subtle and ever recurring interplay of rigorous abstract thought and concrete experience. Turning his attention to the principles of analysis and geometry, he asserts that the "most suggestive and notable achievements of the last century are ... the arithmetical formulation of the concept of the continuum ... and the discovery of non-Euclidean geometry ... ". Then he states Cantor's continuum problem as his first and the consistency issue for the arithmetical axioms as his second problem:

> ... I wish to designate the following as the most important among the numerous questions which can be asked with regard to the axioms: *To prove that they are not contradictory, that is, that a finite number of logical steps based upon them can never lead to contradictory results.*[5]

As to the arithmetical axioms, Hilbert points to his paper *Über den Zahlbegriff* delivered at the Munich meeting of the German Association of Mathematicians in December of 1899. Its title indicates a part of its intellectual context, as Kronecker had published only twelve years earlier a well-known paper with the very same title. In his paper Kronecker sketched a way of introducing irrational numbers, without accepting the general notion. It is precisely to this general concept Hilbert wants to give a proper foundation – using the axiomatic method and following Dedekind.

Hilbert connected the consistency problem also to Cantorian issues. From his correspondence with Cantor we know that he had been aware of Cantor's "inconsistent multiplicities" since 1897. He recognized their effect on Dedekind's logical grounding of analysis and tried to blunt it by a consistency proof for the arithmetical axioms. Such a proof, he stated, provides "the proof of the existence of the totality of real numbers, or – in the terminology of G. Cantor — the proof that

[5] *Ewald*, pp. 1102-03. This is part of the German text from *Hilbert 1901*, p. 55: Vor allem möchte ich unter den zahlreichen Fragen, welche hinsichtlich der Axiome gestellt werden können, dies als das wichtigste Problem bezeichnen, *zu beweisen, daß dieselben unter einander widerspruchslos sind, d.h., daß man auf Grund derselben mittels einer endlichen Anzahl von logischen Schlüssen niemals zu Resultaten gelangen kann, die mit einander in Widerspruch stehen.*

the system of real numbers is a consistent (finished) set."[6] In the Paris Lecture Hilbert re-emphasized and expanded this point by saying,

> In the case before us, where we are concerned with the axioms of real numbers in arithmetic, the proof of the consistency of the axioms is at the same time the proof of the mathematical existence of the complete system ["complete system" is Ewald's translation for "Inbegriff", WS] of real numbers or of the continuum. Indeed, when the proof for the consistency of the axioms shall be fully accomplished, the doubts, which have been expressed occasionally as to the existence of the complete system of real numbers, will become totally groundless.[7]

He proposed then to extend this approach to the Cantorian Alephs. Consequently and more generally, the consistency of appropriate axiom systems was to guarantee the existence of sets or, as he put it in 1905, to provide an *objective criterion* for the consistency of multiplicities.

The background for these foundational issues is explored in Part 2 under the heading *Existential axiomatics*. Part 3 is concerned with *Direct consistency proofs* for equational calculi. Part 4 looks at *Finitist proof theory* and its problems, namely, how first to treat quantifier-free systems of arithmetic and then to deal with quantifiers via the ε-substitution method. Part 5 discusses *Incompleteness* and related mathematical as well as philosophical problems; the essay ends with a very brief outlook beyond 1939.

2 EXISTENTIAL AXIOMATICS

The axiom system Hilbert formulated for the real numbers in 1900 is not presented in the contemporary logical style. Rather, it is given in an algebraic way and assumes that a structure exists whose elements satisfy the axiomatic conditions. Hilbert points out that this *axiomatic* way of proceeding is quite different from the *genetic method* standardly used in arithmetic; it rather parallels the ways of geometry.[8]

[6] *Ewald*, p. 1095. This is part of the German paragraph in *Hilbert 1900*, p. 261: Um die Widerspruchslosigkeit der aufgestellten Axiome zu beweisen, bedarf es nur einer geeigneten Modifikation bekannter Schlußmethoden. In diesem Nachweis erblicke ich zugleich den Beweis für die Existenz des Inbegriffs der reellen Zahlen oder — in der Ausdrucksweise G. Cantors — den Beweis dafür, daß das System der reellen Zahlen eine konsistente (fertige) Menge ist.

[7] *Ewald*, p. 1095. Here is the German text from *Hilbert 1901*, p. 56: In dem vorliegenden Falle, wo es sich um die Axiome der reellen Zahlen in der Arithmetik handelt, ist der Nachweis für die Widerspruchslosigkeit der Axiome zugleich der Beweis für die mathematische Existenz des Inbegriffs der reellen Zahlen oder des Kontinuums. In der That, wenn der Nachweis für die Widerspruchslosigkeit der Axiome völlig gelungen sein wird, so verlieren die Bedenken, welche bisweilen gegen die Existenz des Inbegriffs der reellen Zahlen gemacht worden sind, jede Berechtigung.

[8] In the 1905 Lecture Notes there is a detailed discussion of the genetic foundation (genetische Begründung) with references to textbooks by Lipschitz, Pasch, and Thomae. Hilbert formulates there also the deeply problematic aspects of this method; see the beginning of section 2.2.

> Here [in geometry] one begins customarily by assuming the existence of all the elements, i.e. one postulates at the outset three systems of things (namely, the points, lines, and planes), and then — essentially after the model of Euclid — brings these elements into relationship with one another by means of certain axioms of linking, order, congruence, and continuity.[9]

These geometric ways are taken over for arithmetic or rather, one might argue, are re-introduced into arithmetic by Hilbert; after all, they do have their origin in Dedekind's work on arithmetic and algebra. In the 1920s, Hilbert and Bernays called this methodological approach *existential axiomatics*, because it assumes the existence of suitable systems of objects. The underlying ideas, however, go back in Hilbert's thinking to at least 1893.

Dedekind played a significant role in their evolution not only directly through his foundational essays and his mathematical work, but also indirectly, as his collaborator and friend Heinrich Weber was one of Hilbert's teachers in Königsberg and a marvelous interpreter of Dedekind's perspective. For example, Weber's *Lehrbuch der Algebra* and some of his expository papers are frequently referred to in Hilbert's lectures around the turn from the 19^{th} to the 20^{th} century. It is a slow evolution, which is impeded by the influence of Kronecker's views as to how irrational numbers are to be given. The axiomatic method is to overcome the problematic issues Kronecker saw for the general concept of irrational number. Hilbert's retrospective remarks in *1904** make this quite clear: this general concept created the "greatest difficulties", and Kronecker represented this point of view most sharply. Hilbert seemingly realized only in *1904** that the difficulties are overcome, in a certain way, when the concept of natural number is secured. "The further steps up to the irrational numbers," he claims there, "are then taken in a logically rigorous way and without further difficulties."[10] Let us look at the details of this remarkable evolution, as far as it can be made out from the available documents.

[9] *Ewald*, p. 1092. Here is the German text from *Hilbert 1900*, p. 257: Hier [in der Geometrie] pflegt man mit der Annahme der Existenz von sämtlichen Elementen zu beginnen, d.h. man setzt von vornherein drei Systeme von Dingen, nämlich die Punkte, die Geraden und die Ebenen, und bringt sodann diese Elemente — wesentlich nach dem Vorbilde von Euklid — durch gewisse Axiome, nämlich die Axiome der Verknüpfung, der Anordnung, der Kongruenz und der Stetigkeit, miteinander in Beziehung. Hilbert obviously forgot to mention, between the axioms of order and of congruence, the axiom of parallels.

[10] These matters are discussed in *Hilbert 1904**, pp. 164-167. One most interesting part consists of these remarks on pp. 165-166: Die Untersuchungen in dieser Richtung [foundations for the real numbers] nahmen lange Zeit den breitesten Raum ein. Man kann den Standpunkt, von dem dieselben ausgingen, folgendermaßen charakterisieren: Die Gesetze der ganzen Zahlen, der Anzahlen, nimmt man vorweg, begründet sie nicht mehr; die Hauptschwiergkeit wird in jenen Erweiterungen des Zahlbegriffs (irrationale und weiterhin komplexe Zahlen) gesehen. Am schärfsten wurde dieser Standpunkt von Kronecker vertreten. Dieser stellte geradezu die Forderung auf: Wir müssen in der Mathematik jede Tatsache, so verwickelt sie auch sein möge, auf Beziehungen zwischen ganzen rationalen Zahlen zurückführen; die Gesetze dieser Zahlen andrerseits müssen wir ohne weiteres hinnehmen. Kronecker sah in den Definitionen der irrationalen Zahlen Schwierigkeiten und ging soweit, dieselben garnicht anzuerkennen.

2.1 Axioms for continuous systems

Hilbert introduces the axioms for the real numbers in *1900* as follows: "We think of a system of things, and we call them numbers and denote them by $a, b, c \ldots$ We think of these numbers in certain mutual relations, whose precise and complete description is obtained through the following axioms."[11] Then the axioms for an ordered field are formulated and rounded out by the requirement of continuity via the Archimedean axiom and the axiom of completeness (Axiom der Vollständigkeit). This formulation is not only in the spirit of the geometric ways, but actually mimics the contemporaneous and paradigmatically modern axiomatic presentation of *Grundlagen der Geometrie*.

> We think of three different systems of things: we call the things of the first system *points* and denote them by $A, B, C, \ldots$; we call the things of the second system *lines* and denote them by $a, b, c, \ldots$; we call the things of the third system *planes* and denote them by $\alpha, \beta, \gamma, \ldots; \ldots$ We think of the points, lines, planes in certain mutual relations $\ldots$; the precise and complete description of these relations is obtained by the *axioms of geometry*.[12]

Five groups of geometric axioms follow and, in the first German edition, the fifth group consists of just the Archimedean axiom. In the French edition of 1900 and the second German edition of 1903, the completeness axiom is included. The latter axiom requires in both the geometric and the arithmetic case that the assumed structure is maximal, i.e., any extension satisfying all the remaining axioms must already be contained in it. This formulation is frequently criticized as being metamathematical and, to boot, of a peculiar sort.[13] However, it is just an ordinary mathematical formulation, if the algebraic character of the axiom system is kept in mind. In the case of arithmetic we can proceed as follows: call a system A *continuous* when its elements satisfy the axioms of an ordered field and the Archimedean axiom, and call it *fully continuous* if and only if A is continuous and for any system B, if $A \subseteq B$ and B is continuous, then $B \subseteq A$.[14] Hilbert's arithmetic axioms

[11] *Hilbert 1900*, pp. 257-258. The German text is: Wir denken ein System von Dingen; wir nennen diese Dinge Zahlen und bezeichnen sie mit a, b, c, ... Wir denken diese Zahlen in gewissen gegenseitigen Beziehungen, deren genaue und vollständige Beschreibung durch die folgenden Axiome geschieht.

[12] *Hilbert 1899*, p. 4; in *LFMP1*, p. 437. The German text is: Wir denken drei verschiedene Systeme von Dingen: die Dinge des ersten Systems nennen wir *Punkte* und bezeichnen sie mit $A, B, C, \ldots$; die Dinge des zweiten Systems nennen wir *Gerade* und bezeichnen sie mit $a, b, c, \ldots$; die Dinge des dritten Systems nennen wir *Ebenen* und bezeichnen sie mit $\alpha, \beta, \gamma, \ldots; \ldots$ Wir denken die Punkte, Geraden, Ebenen in gewissen gegenseitigen Beziehungen $\ldots$; die genaue und vollständige Beschreibung dieser Beziehungen erfolgt durch die Axiome der Geometrie.

[13] In *Torretti*, for example, one finds on p. 234 this observation: "... the completeness axiom is what nowadays one would call a metamathematical statement, though one of a rather peculiar sort, since the theory with which it is concerned includes Axiom V2 [i.e., the completeness axiom, WS] itself."

[14] This is dual to Dedekind's definition of simply infinite systems: here we are focusing on maximal extensions, and there on minimal substructures. Indeed, the formulation in the above text is to be taken in terms of substructures in the sense of contemporary model theory.

characterize the fully continuous systems uniquely up to isomorphism.

Hilbert thought about axiom systems in this existential way already in *Die Grundlagen der Geometrie*, his first lectures on the *foundations* of geometry. He had planned to give them in the summer term of 1893, but their presentation was actually shifted to the following summer term. Using the notions "System" and "Ding" so prominent in Dedekind's *1888*, he formulated the central question as follows:

> What are the necessary and sufficient and mutually independent conditions a system of things has to satisfy, so that to each property of these things a geometric fact corresponds and conversely, thereby making it possible to completely describe and order all geometric facts by means of the above system of things.[15]

At some later point, Hilbert inserted the remark that this axiomatically characterized system of things is a "complete and simple image of geometric reality."[16] The image of "image" is attributed on the next page to Hertz and extended to cover also logical consequences of the images; they are to coincide with images of the consequences: "The axioms, as Hertz would say, are images or symbols in our mind, such that consequences of the images are again images of the consequences, i.e., what we derive logically from the images is true again in nature." In the Introduction to the Notes for the 1898-99 lectures *Elemente der Euklidischen Geometrie* written by Hilbert's student Hans von Schaper, this central question is explicitly connected with Hertz's *1894*:

> Using an expression of Hertz (in the introduction to the *Prinzipien der Mechanik*) we can formulate our main question as follows: What are the necessary and sufficient and mutually independent conditions a system of things has to be subjected to, so that to each property of these things a geometric fact corresponds and conversely, thereby having these things provide a complete "image" of geometric reality.[17]

[15] *Toepell*, pp. 58-59 and *LFMP1*, pp. 72-73. The German text is: Welches sind die nothwendigen und hinreichenden und unter sich unabhängigen Bedingungen, die man an ein System von Dingen stellen muss, damit jeder Eigenschaft dieser Dinge eine geometrische Thatsache entspricht und umgekehrt, so dass also mittelst obigen Systems von Dingen ein vollständiges Beschreiben und Ordnen aller geometrischen Thatsachen möglich ist.

[16] In *Toepell*, p. 59 and *LFMP1*, pp. 73 and 74. The German text is: ... so dass also unser System ein vollständiges und einfaches Bild der geometrischen Wirklichkeit werde. – The next German text is: Die Axiome sind, wie Hertz sagen würde, Bilder oder Symbole in unserem Geiste, so dass Folgen der Bilder wieder Bilder der Folgen sind, d.h. was wir aus den Bildern logisch ableiten, stimmt wieder in der Natur. On the very page of the last quotation, in note 7 of *LFMP1*, the corresponding and much clearer remark from *Hertz 1894* is quoted: Wir machen uns innere Scheinbilder oder Symbole der äußeren Gegenstände, und zwar machen wir sie von solcher Art, daß die denknotwendigen Folgen der Bilder stets wieder die Bilder seien von naturnotwendigen Folgen der abgebildeten Gegenstände.

[17] *Toepell*, p. 204 and *LFMP1*, p. 303. The German text is: Mit Benutzung eines Ausdrucks von Hertz (in der Einleitung zu den *Prinzipien der Mechanik*) können wir unsere Hauptfrage so formulieren: Welches sind die nothwendigen und hinreichenden und unter sich unabhängigen

One can see in these remarks the shape of a certain logical or set theoretic structuralism (that is *not* to be equated with contemporary philosophical forms of such a position) in the foundations of mathematics and of physics; geometry is viewed as the most perfect natural science.[18] How simple or how complex are these systems or structures supposed to be? What characteristics must they have in order to be adequate for both the systematic internal development of the corresponding theory and external applications? What are the things whose system is implicitly postulated? What is the mathematical connection, in particular, between the arithmetic and geometric structures?

Dedekind gives in *Stetigkeit und irrationale Zahlen* a precise notion of continuity for analysis; however, before doing so he shifts his considerations to the geometric line. He observes that every point effects a cut, but — and that he views as central for the *continuity* or *completeness* of the line — every cut is effected by a point.[19] Transfering these reflections to cuts of rational numbers and, of course, having established that there are infinitely many cuts of rationals that are not effected by rationals, he "creates" a new, an irrational number α for any cut not effected by a rational: "We will say that this number α corresponds to that cut or that it effects this cut."[20] Dedekind then introduces an appropriate ordering between cuts and establishes that the system of all cuts is indeed continuous. He shows that the continuity principle has to be accepted, if one wants to have the least upper bound principle for monotonically increasing, but bounded sequences of reals, as the two principles are indeed equivalent. He writes:

> This theorem is equivalent to the principle of continuity, i.e., it loses its validity as soon as even just a single real number is viewed as not being contained in the domain R; ...[21]

As it happens, Hilbert's first formulation of a geometric continuity principle in *1894** (p. 92 of *LFMP1*) is a least upper bound principle for bounded sequences of geometric points.

Dedekind emphasizes in *1872* the necessity of rigorous arithmetic concepts for analyzing the notion of space. He points out, reflecting the view of Gauss concerning the fundamental difference between arithmetic and geometry, that space

Bedingungen, denen man ein System von Dingen unterwerfen muss, damit jeder Eigenschaft dieser Dinge eine geometrische Thatsache entspreche und umgekehrt, damit also diese Dinge ein vollständiges und einfaches "Bild" der geometrischen Wirklichkeit seien.

[18] There is a complementary discussion on pp. 103-104 and 119-122 of *LFMP1*. Hilbert's perspective is echoed in the views of Ernest Nagel and Patrick Suppes; cf. also van Fraassen, *The Scientific Image*.

[19] Dedekind uses *Stetigkeit* and *Vollständigkeit* interchangeably. The reader should notice the freedom with which Dedekind reinterprets in the following considerations an axiomatically characterized concept, namely a certain order relation: it is interpreted for the rational numbers, the geometric line, and the system of all cuts of rationals.

[20] *Dedekind 1872*, p. 13.

[21] *Dedekind* 1872, p. 20. The German text is: Dieser Satz ist äquivalent mit dem Prinzip der Stetigkeit, d.h. er verliert seine Gültigkeit, sobald man auch nur eine einzige reelle Zahl in dem Gebiete R als nicht vorhanden ansieht; ...

need not be continuous.[22] The required arithmetization of analysis is established in Dedekind's thinking about the matter through three steps, namely, i) a logical characterization of the natural numbers, ii) the extension of that number concept to integers and rational numbers, and iii) the further extension to all real numbers we just discussed. The first objective is accomplished in *1888* via simply infinite systems, a notion that emerged in the 1870s after Dedekind had finished his essay *1872*. (See *Sieg & Schlimm*; in that paper one finds also a discussion of Dedekind's approach to the second issue based on unpublished manuscripts.) Dedekind gives his final characterization in article 71 of *1888* in the existential, algebraic form that is used later by Hilbert. But he avoids self-consciously the term axiom and speaks only of conditions (Bedingungen), which are easily seen to amount in this case to the well-known Peano axioms.

Dedekind sought the ultimate basis for arithmetic in logic, broadly conceived. That view is formulated in the preface to *1888*:

> In calling arithmetic (algebra, analysis) merely a part of logic, I already claim that I consider the number concept as entirely independent of notions or intuitions of space and time, that I consider it rather as an immediate product of the pure laws of thought.[23]

Important for our considerations is the fact that Dedekind had developed towards the end of the 1870s not only the general concept of "system", but also that of a "mapping between systems". The latter notion corresponds for him to a crucial capacity of our mind (pp. III-IV of the preface to *1888*). He uses these concepts to formulate and prove that all simply infinite systems are isomorphic. He concludes from this result, in contemporary model theoretic terms, that simply infinite systems are all elementarily equivalent and uses that fact to characterize the science of arithmetic:

> The relations or laws, which are solely derived from the conditions $\alpha, \beta, \gamma, \delta$ in 71 and are therefore in all ordered systems always the same ... form the next object of the *science of numbers* or of *arithmetic*.[24]

Natural numbers are not viewed as objects of a particular system, but are obtained by free creation or what W.W. Tait has called "Dedekind abstraction"; that is clearly articulated in article 73 of *1888*.

Similar metamathematical investigations are adumbrated for real numbers in *Stetigkeit und irrationale Zahlen*. Dedekind can be interpreted, in particular, as

[22] In *Dedekind 1888*, p. VII, a particular countable model of reals is described in which all of Euclid's construction can be interpreted.

[23] *Dedekind 1888*, p. III. The German text is: Indem ich die Arithmetik (Algebra, Analysis) nur einen Teil der Logik nenne, spreche ich schon aus, daß ich den Zahlbegriff für gänzlich unabhängig von den Vorstellungen oder Anschauungen des Raumes und der Zeit, daß ich ihn vielmehr für einen unmittelbaren Ausfluß der reinen Denkgesetze halte.

[24] *Dedekind 1888*, p. 17. The German text is: Die Beziehungen oder Gesetze, welche ganz allein aus den Bedingungen $\alpha, \beta, \gamma, \delta$ in 71 abgeleitet werden und deshalb in allen geordneten Systemen immer dieselben sind ... bilden den nächsten Gegenstand der *Wissenschaft von den Zahlen* oder der *Arithmetik*.

having characterized *fully continuous systems* as ordered fields that satisfy his principle of continuity, i.e., the principle that every cut of such a system is effected by an element. The system of all cuts of rational numbers constitutes a model; that is in modern terms the essential content of sections 5 and 6 of *1872*, where Dedekind verifies that this system satisfies the continuity principle. In contemporaneous letters to Lipschitz, Dedekind emphasizes that real numbers should not be identified with cuts, but rather viewed as newly created things corresponding to cuts.

The informal comparison of the geometric line with the system of cuts of rational numbers in sections 2 and 3 of *1872* contains almost all the ingredients for establishing rigorously that these two structures are isomorphic; missing is the concept of mapping. As I mentioned, that concept was available to Dedekind in 1879 and with it these considerations can be extended directly to show that arbitrary fully continuous systems are isomorphic. Finally, the methodological remarks about the arithmetic of natural numbers can now be extended to that of the real numbers. As a matter of fact, Hilbert articulates Dedekind's way of thinking of the system of real numbers in his *1922*, when describing the *axiomatische Begründungsmethode* for analysis:

> The continuum of real numbers is a system of things that are connected to each other by certain relations, so-called axioms. In particular the definition of the real numbers by the Dedekind cut is replaced by two continuity axioms, namely, the Archimedian axiom and the so-called completeness axiom. In fact, Dedekind cuts can then serve to determine the individual real numbers, but they do not serve to define [the concept of] real number. On the contrary, conceptually a real number is just a thing of our system. ... The standpoint just described is altogether logically completely impeccable, and it only remains thereby undecided, whether a system of the required kind can be thought, i.e., whether the axioms do not lead to a contradiction.[25]

That is fully in Dedekind's spirit: Hilbert's critical remarks about the definition of real numbers as cuts do not apply to Dedekind, as should be clear from the above discussion, and the issue of consistency had been an explicit part of Dedekind's logicist program.

[25] *Hilbert 1922*, pp. 158-159. The German text is: Das Kontinuum der reellen Zahlen ist ein System von Dingen, die durch bestimmte Beziehungen, sogenannte Axiome, miteinander verknüpft sind. Insbesondere treten an Stelle der Definition der reellen Zahlen durch den Dedekindschen Schnitt die zwei Stetigkeitsaxiome, nämlich das Archimedische Axiom und das sogenannte Vollständigkeitsaxiom. Die Dedekindschen Schnitte können dann zwar zur Festlegung der einzelnen reellen Zahlen dienen, aber sie dienen nicht zur Definition der reellen Zahl. Vielmehr ist begrifflich eine reelle Zahl eben ein Ding unseres Systems. ... Der geschilderte Standpunkt ist vollends logisch vollkommen einwandfrei, und es bleibt nur dabei unentschieden, ob ein System der verlangten Art denkbar ist, d.h. ob die Axiome nicht etwa auf einen Widerspruch führen.

2.2 Consistency and logical models

Logicians in the 19^{th} century viewed consistency of a notion from a semantic perspective. In that tradition, Dedekind addressed the problem for his characterization of natural numbers via simply infinite systems. The methodological need for doing that is implicit in *1872*, but is formulated most clearly in a letter to Keferstein written in 1889, a year after the publication of *Was sind und was sollen die Zahlen?*

> After the essential nature of the simply infinite system, whose abstract type is the number sequence N, had been recognized in my analysis (71, 73) the question arose: does such a system *exist* at all in the realm of our thoughts? Without a logical proof of existence, it would always remain doubtful whether the notion of such a system might not perhaps contain internal contradictions. Hence the need for such a proof (articles 66 and 72 of my essay).[26]

Dedekind attempted in article 66 to prove the existence of an infinite system and to provide in article 72 an example of a simply infinite system within logic. In other words, he attempted to establish the existence of a *logical* model and to guarantee in this way that the notion of a simply infinite system indeed does not contain internal contradictions.

Hilbert turned to considerations for natural numbers only around 1904. Until then he had taken for granted their proper foundation and focused on the theory of real numbers: as I reported earlier from his *1904**, Hilbert thought that the concept of irrational numbers created the greatest difficulties. These difficulties have been overcome, Hilbert claims in *1904**. When the concept of natural number is secured, he continues there, the further steps towards the real numbers can be taken without a problem. This dramatic change of viewpoint helps us to understand more clearly some aspects of Hilbert's earlier considerations related to the consistency of the arithmetic of real numbers. What did Hilbert see as the "greatest difficulties" for the general concept of irrational numbers? That question will be explored now.

In contrast to Dedekind, Hilbert had formulated a quasi-syntactic notion of consistency in *Über den Zahlbegriff* and in *Grundlagen der Geometrie*, namely, no finite number of logical steps leads from the axioms to a contradiction. This notion is "*quasi*-syntactic", as no deductive principles are explicitly provided. Hilbert did not seek to prove consistency by syntactic methods. On the contrary, the relative consistency proofs given in *Grundlagen der Geometrie* are all straightforwardly semantic. To prove the consistency of arithmetic one needs, so Hilbert in *Über*

[26] In *van Heijenoort*, p. 101. The German text is: Nachdem in meiner Analyse der wesentliche Charakter des einfach unendlichen Systems, deßen abstracter Typus die Zahlenreihe N ist, erkannt war (71, 73), fragte es sich: *existirt* überhaupt ein solches System in unserer Gedankenwelt? Ohne den logischen Existenz-Beweis würde es immer zweifelhaft bleiben, ob nicht der Begriff eines solchen Systems vielleicht innere Widersprüche enthält. Daher die Nothwendigkeit solcher Beweise (66, 72 meiner Schrift).

den Zahlbegriff, "a suitable modification of familiar methods of reasoning." In the Paris Lecture Hilbert suggests finding a *direct* proof and remarks more expansively:

> I am convinced that it must be possible to find a direct proof for the consistency of the arithmetical axioms [as proposed in *Über den Zahlbegriff* for the real numbers; WS], by means of a careful study and suitable modification of the known methods of reasoning in the theory of irrational numbers.[27]

Hilbert believed at this point, it seems to me, that the genetic build-up of the real numbers could *somehow* be exploited to yield the blueprint for a consistency proof in Dedekind's logicist style. There are difficulties with the genetic method that prevent it from providing directly a proper foundation for the general concept of irrational numbers. Hilbert's concerns are formulated most clearly in these retrospective remarks from *1905**:

> It [the genetic method, WS] defines things by generative processes, not by properties what really must appear to be desirable. Even if there is no ojection to defining fractions as systems of two integers, the definition of irrational numbers as a sytem of infinitely many numbers must appear to be dubious. Must this number sequence be subject to a law, and what is to be understood by a law? Is an irrational number being defined, if one determines a number sequence by throwing dice? These are the kinds of questions with which the genetic perspective has to be confronted.[28]

Precisely this issue was to be overcome (or to be sidestepped) by the axiomatic method. In *Über den Zahlbegriff* Hilbert writes:

> Under the conception described above, the doubts which have been raised against the existence of the totality of real numbers (and against the existence of infinite sets generally) lose all justification; for by the set of real numbers we do not have to imagine, say, the totality of all possible laws according to which the elements of a fundamental

[27] *Ewald*, p. 1104. The German text from *Hilbert 1901*, p. 55 is: Ich bin nun überzeugt, daß es gelingen muß, einen direkten Beweis für die Widerspruchslosigkeit der arithmetischen Axiome zu finden, wenn man die bekannten Schlußmethoden in der Theorie der Irrationalzahlen im Hinblick auf das bezeichnete Ziel genau durcharbeitet und in geeigneter Weise modifiziert. In *Bernays 1935*, pp. 198-9, this idea is reemphasized: Zur Durchführung des Nachweises gedachte Hilbert mit einer geeigneten Modifikation der in der Theorie der reellen Zahlen angewandten Methoden auszukommen.

[28] *Hilbert 1905**, pp. 10-11. The German text is: Sie definiert Dinge durch Erzeugungsprozesse, nicht durch Eigenschaften, was doch eigentlich wünschenswert erscheinen muß. Ist nun auch gegen die Definition der Brüche als System zweier Zahlen nichts einzuwenden, so muß doch die Definition der Irrationalzahlen als ein System von unendlich vielen Zahlen bedenklich erscheinen. Muß diese Zahlenreihe einem Gesetz unterliegen, und was hat man unter einem Gesetz zu verstehen? Wird auch eine Irrationalzahl definiert, wenn man eine Zahlenreihe durch Würfeln feststellt? Dieser Art sind die Fragen, die man der genetischen Betrachtung entgegenhalten muß.

> sequence can proceed, but rather — as just described — a system of things whose mutual relations are given by the finite and closed system of axioms I–IV, ...[29]

He articulates in the *Paris Lecture* exactly the same point by re-emphasizing "the continuum ... is not the totality of all possible series in decimal fractions, or of all possible laws according to which the elements of a fundamental sequence may proceed." Rather, it is a system of things whose mutual relations are governed by the axioms. The completeness axiom can be plausibly interpreted as guaranteeing the continuity of the system without depending on any particular method of generating real numbers.

Going back to the 1889–90 lectures *Einführung in das Studium der Mathematik*, one sees that Hilbert then shared a Kroneckerian perspective on irrationals. He starts out with a brief summary of "Das Nothwendigste über den Zahlbegriff" and uses Dedekind cuts to define the notion of irrational number. He shifts for the practical presentation of real numbers from cuts to fundamental sequences and asserts, making the problematic aspect fully explicit, "It is absolutely necessary to have a general method for the representation of irrational numbers."[30] The need to have a method of representation is not bound to fundamental sequences. Cuts are to be given by a *rule* that allows the partition of all rationals into smaller and greater ones: the real number effecting the cut is taken to be "the carrier of that rule". (This perspective is re-iterated in the later lectures, in particular, in *1899–1900**, pp. 10-11.) So we have a sense of the difficulties as Hilbert saw them. In order to support the claim that Hilbert thought, nevertheless, that the genetic build-up of the real numbers could *in some way* be exploited to yield a logicist consistency proof, we look closely at Hilbert's treatment of arithmetic in later lectures, in particular, *1894*, 1897–98**, and *1899–1900**.

Let me begin with some deeply programmatic, logicist statements found in the *Introduction* to the notes for *Elemente der Euklidischen Geometrie* presented in 1898-99. Hilbert maintains there, "It is important to fix precisely the starting-point of our investigations: as given we consider the laws of pure logic and in particular all of arithmetic."[31] He adds parenthetically, "On the relation between

[29] *Ewald*, p. 1095. Here is the German text from *Hilbert 1900*, p. 261: Die Bedenken, welche gegen die Existenz des Inbegriffs aller reellen Zahlen und unendlicher Mengen überhaupt geltend gemacht worden sind, verlieren bei der oben gekennzeichneten Auffassung jede Berechtigung: unter der Menge der reellen Zahlen haben wir uns hiernach nicht etwa die Gesamtheit aller möglichen Gesetze zu denken, nach denen die Elemente einer Fundamentalreihe fortschreiten können, sondern vielmehr – wie eben dargelegt ist – ein System von Dingen, deren gegenseitige Beziehungen durch das obige endliche und abgeschlossene System von Axiomen I-IV gegeben sind, ...

[30] *Hilbert 1889–90**, p. 3. The German text is: Es ist unumgänglich nöthig, eine allgemeine Methode zur Darstellung der irrationalen Zahlen zu haben.

[31] Clearly, arithmetic is here taken in the narrower sense of the theory of natural numbers. The German text is found in *Toepell*, pp. 203-4, and in *LFMP1*, p. 303: Es ist von Wichtigkeit, den Ausgangspunkt unserer Untersuchungen genau zu fixieren: Als gegeben betrachten wir die Gesetze der reinen Logik und speciell die ganze Arithmetik. (Ueber das Verhältnis zwischen Logik und Arithmetik vgl. Dedekind, Was sind und was sollen die Zahlen?) Unsere Frage wird

logic and arithmetic cf. Dedekind, *Was sind und was sollen die Zahlen?*" Clearly, for Dedekind arithmetic is a part of logic, and for Hilbert in 1899 that is not a novel insight. On the contrary: Hilbert opens the lectures *1889–90* with the sentence, "The theoretical foundation for the mathematical sciences is the concept of positive integer, defined as *Anzahl*." This remark is deepened in the long introduction to the lectures on *Projektive Geometrie* given in 1891, where Hilbert emphasizes that the results of pure mathematics are obtained by pure thought (reines Denken). He lists as parts of pure mathematics number theory, algebra, and complex analysis (Funktionentheorie). Mirroring Dedekind's perspective on the arithmetization of analysis from the preface to *1888* (p. VI), Hilbert asserts that the results of pure mathematics *can* and *have to be reduced* to relations between integers:

> *Nowadays* a proposition is being considered as proved only, when it *ultimately expresses* a relation between integers. Thus, the *integer is the element*. We can *obtain by pure thinking* the concept of integer, for example <by> *counting thoughts themselves*. Methods, foundations of pure mathematics belong to the sphere of pure thought.[32]

This very paragraph ends with the striking assertion, "I don't need anything but purely logical thought, when dealing with number theory or algebra." Hilbert expresses here very clearly a thoroughgoing Dedekindian logicism with respect to natural numbers.[33]

Die Grundlagen der Geometrie from 1894 have a special section on "Die Einführung der Zahl". After all, Hilbert argues, "In all of the exact sciences one obtains precise results only after the introduction of number. It is always of deep epistemological significance to see in detail, how this measuring is done."[34] The number concept discussed now is that of real numbers; it is absolutely central for the problems surrounding the quadrature of the circle treated in the winter term 1894-95. The quadrature problem is the red thread for historically informed lectures entitled, not surprisingly, *Quadratur des Kreises*. Its solution requires the subtlest mathematical speculations with considerations from geometry as well as from analysis and number theory. In the winter term 1897-98 Hilbert presents these topics again, but this time in the *second* part of the course; its new *first* part is devoted

dann sein: Welche Sätze müssen wir zu dem eben definierten Bereich "adjungieren", um die Euklidische Geometrie zu erhalten? In *Toepell* one finds additional remarks of interest for the issues discussed here, namely, on pp. 195, 206, and 227-8.

[32] *LFMP1*, p. 22. The German text is: *Es gilt heutzutage* ein Satz erst dann als bewiesen, wenn er eine Beziehung zwischen ganzen Zahlen *in letzter Instanz zum Ausdruck bringt*. Also die *ganze Zahl ist das Element*. Zum Begriff der ganzen Zahl können wir auch *durch reines Denken gelangen*, etwa <indem ich> die *Gedanken selber zähle*. *Methoden, Grundlagen der reinen Mathematik* gehören dem reinen Denken an.

[33] L.c., p.22. The German sentence is: Ich brauche weiter nichts als *rein logisches* Denken, wenn ich mit Zahlentheorie oder Algebra mich beschäftige. In his extensive editorial note 6 on p. 23 of *LFMP1*, Haubrich also points out the deep connection to Dedekind's views.

[34] L.c., p. 85. The German text is: In allen exakten Wissenschaften gewinnt man erst dann präzise Resultate, wenn die Zahl eingeführt ist. Es ist stets von hoher erkenntnistheoretischer Bedeutung zu verfolgen, wie dies Messen geschieht.

entirely to a detailed discussion of the concept of number. That expansion is indicated by the new title of the lectures, *Zahlbegriff und Quadratur des Kreises*.

Hilbert remarks that the foundations of analysis have been examined through the work of Weierstrass, Cantor, and Dedekind. He lists three important points:

> Today [we take as a] principle: Reduction to the laws for the whole rational numbers and purely logical operations.
>
> 1, 2, 3 ... = Anzahl = natural number
>
> How one gets from this number [concept] to the most general concept of complex number, that is the foundation, i.e., the theory of the concept of number.[35]

In the lectures *1897–98** the genetic build-up is to be captured axiomatically; that I take to be the gist of the discussion of the third point in the above list, namely, how one gets from the natural numbers to the most general concept of complex number. The motivating focus lies on the solution of problems by an extension of the axioms (Lösung von Problemen durch Erweiterung der Axiome); the discussion ends with the remark, "We have extended the number concept in such a way, or to put it in a better way, we have enlarged the stock of axioms in such a way, that we can solve all definite equations."[36] At each step in this extending process consistency of the axioms is required, but Hilbert does not give arithmetical or logical proofs, only geometric pictures.

In the winter term 1899-1900, Hilbert gave again the lectures *Zahlbegriff und Quadratur des Kreises* and refined the discussion of the axioms for the reals significantly; it is a transition to his paper *Über den Zahlbegriff*. Pages 38-43 are added to the notes from the previous lecture, and Hilbert extends on them the discussion of "komplexe Zahlensysteme" from *Grundlagen der Geometrie*.[37] He formulates the continuity principle as "Bolzano's Axiom", i.e., every fundamental sequence has a limit; he shows also that this axiom has the Archimedian one as a consequence. Strikingly, he actually claims (on p. 42) that the full system is consistent! Before completing his *1900*, Hilbert obviously recognized i) that he could formulate the second continuity principle in such a way that it did not imply the Archimedian axiom, and ii) that the consistency of the system had *not* been established. So we are back to the issue from the very beginning of this subsection. The long digression does not provide a clear answer, but a seemingly rhetorical question: How

[35] *Hilbert 1897–98**, p. 2. The German text is: Heute Grundsatz: Zurückführung auf die Gesetze der ganzen rationalen Zahlen und rein logische Operationen.

1, 2, 3 ... = Anzahl = natürliche Zahl

Wie man von dieser Zahl zum allgemeinsten Begriff der komplexen Zahl kommt, das ist jenes Fundament, d.h. die Lehre vom Zahlbegriff.

[36] *Hilbert 1897-98**, p. 23. The German text is: Wir haben den Zahlbegriff so erweitert, oder besser gesagt, den Bestand der Axiome so vermehrt, dass wir alle definiten Gleichungen lösen können. Hallett gives a complementary account of the completeness principle that emphasizes its role in Hilbert's work concerning geometry; that is presented in section 5 of his Introductory Note to *Hilbert 1899*, pp. 426-435 of *LFMP1*.

[37] Indeed, pages 38 and 39 must be the page proofs of the *Festschrift*, pp. 26-28.

could Hilbert think about addressing the consistency problem by "a careful study and suitable modification of the known methods of resoning in the theory of irrational numbers," if he did not have in mind a programmatically guided refinement of the steps presented here, i.e., of the foundation they constitute?

2.3 Consistency and proofs

Hilbert knew about the difficulties in set theory through his correpondence with Cantor, but he did not — as far as I can see — move away from this programmatic position and the associated strategy for proving consistency until 1903 or, at the latest, 1904. One reason for making this conjecture rests on the fact that Hilbert gave his lectures on *Zahlbegriff und Quadratur des Kreises* again in 1901-02; we have only the "Disposition" of those lectures, but it is very detailed and does not contain any indication whatsoever of substantive changes. In the summer term of 1904, Hilbert lectured again on *Zahlbegriff und Quadratur des Kreises*. In the notes written by Max Born one finds a dramatic change: Hilbert discusses for the first time the paradoxes in detail and sketches various foundational approaches.[38] These discussions are taken up in Hilbert's talk at the Heidelberg Congress in August of that year, where he presents a syntactic approach to the consistency problem. The goal is still to guarantee the existence of a suitable system, and the method seems inspired by one important aspect of the earlier investigations, as the crucial new step is the *simultaneous development of arithmetic and logic*!

Hilbert's view of the geometric axioms as characterizing a system of things that presents a "complete and simple image of geometric reality" is, after all, complemented by a traditional one: the axioms must allow to establish, purely logically, all geometric facts and laws. This role of the axioms, though clearly implicit in the formulation of the central question of *1894*, is emphasized in the notes *Elemente der Euklidischen Geometrie* and described in the Introduction to the *Festschrift* in a methodologically refined way:

> The present investigation is a new attempt of formulating for geometry a *simple* and *complete* system of mutually independent axioms; it is also an attempt of deriving from them the most important geometric propositions in such a way that the significance of the different groups of axioms and the import of the consequences of the individual axioms is brought to light as clearly as possible.[39]

[38] Peckhaus characterizes 1903 in his book *1990* as a "Zäsurjahr" in Göttingen; see pp. 55–58. Peckhaus also points out that Dedekind, in that very year, did not allow the republication of his *1888*, because — as Dedekind put it in the preface to the third edition of 1911 — he had doubts concerning some of the important foundations of his view. Further evidence for this conjecture is presented at the beginning of section 3 below, in particular Hilbert's own recollections and remarks by Bernays.

[39] *Hilbert 1899*, Einleitung, p. 3, and *LFMP1*, p. 436. The German text is: Die vorliegende Untersuchung ist ein neuer Versuch, für die Geometrie ein *einfaches* und *vollständiges* System von einander *unabhängiger* Axiome aufzustellen und aus denselben die wichtigsten geometrischen Sätze in der Weise abzuleiten, dass dabei die Bedeutung der verschiedenen Axiomgruppen und

This second important role of axioms is reflected also in the Paris Lecture, where Hilbert states that the totality of real numbers is "... a system of things whose mutual relations are governed by the axioms set up and for which all propositions, and only those, are true which can be derived from the axioms by a finite number of logical inferences." As in Hilbert's opinion "the concept of the continuum is strictly logically tenable in this sense only," the axiomatic method has to confront two fundamental problems that are formulated in *1900* at first for geometry and then also for arithmetic:

> The necessary task then arises of showing the consistency and the completeness of these axioms, i.e., it must be proved that the application of the given axioms can never lead to contradictions, and, further, that the system of axioms suffices to prove all geometric propositions.[40]

It is not clear, whether completeness of the respective axioms requires the proof of *all* true geometric or arithmetic statements, or whether provability of those that are part of the established corpora is sufficient; the latter would be a *quasi-empirical* notion of completeness. It should be mentioned that Dedekind attended in his *1888* consciously to the quasi-empirical completeness of the axiomatic conditions characterizing simply infinite systems. Through the proof principle of induction and the definition principle of recursion he provided the basic principles needed for the actual development of elementary number theory. The same observation can be made for the principle of continuity and Dedekind's treatment of analysis in his *1872*.

But what are the logical steps that are admitted in proofs? Frege criticized Dedekind on that point in the Preface of his *Grundgesetze der Arithmetik*, claiming that the brevity of Dedekind's development of arithmetic in *1888* is only possible, "because much of it is not really proved at all. " Frege continues:

> ... nowhere is there a statement of the logical or other laws on which he builds, and, even if there were, we could not possibly find out whether really no others were used – for to make that possible the proof must be not merely indicated but completely carried out.[41]

Apart from making the logical principles explicit there is an additional aspect that is hinted at in Frege's critique (and detailed in other writings). Though Frege's critique applies as well to Hilbert's *Grundlagen der Geometrie*, Poincaré's 1902-review of that book brings out this additional aspect in a dramatic way, namely, the idea of formalization as machine executability. Poincaré writes: "M.

die Tragweite der aus den einzelnen Axiomen zu ziehenden Folgerungen möglichst klar zu Tage tritt.

[40] *Ewald*, pp. 1092-3. The German text from *Hilbert 1900* is: Es entsteht dann die notwendige Aufgabe, die Widerspruchslosigkeit und Vollständigkeit dieser Axiome zu zeigen, d.h. es muß bewiesen werden, daß die Anwendung der aufgestellten Axiome nie zu Widersprüchen führen kann, und ferner, daß das System der Axiome zum Nachweis aller geometrischen Sätze ausreicht.

[41] *Grundgestze der Arithmetik*, p. 139 of *Translations from the philosophical writings of Gottlob Frege*, Peter Geach and Max Black (eds.), Oxford 1977.

Hilbert has tried, so-to-speak, putting the axioms in such a form that they could be applied by someone who doesn't understand their meaning, because he has not ever seen either a point, or a line, or a plane. It must be possible, according to him [Hilbert, WS], to reduce reasoning to purely mechanical rules." Indeed, Poincaré suggests giving the axioms to a reasoning machine, like Jevons' logical piano, and observing whether all of geometry would be obtained. Such formalization might seem "artificial and childish", were it not for the important question of completeness:

> Is the list of axioms complete, or have some of them escaped us, namely those we use unconsciously? ... One has to find out whether geometry is a logical consequence of the explicitly stated axioms, or in other words, whether the axioms, when given to the reasoning machine, will make it possible to obtain the sequence of all theorems as- output [of the machine, WS].[42]

At issue is, of course, what logical steps can the reasoning machine take? In Hilbert's lectures one finds increasingly references to, and investigations of, logical calculi. As to the character of these calculi, see Peckhaus' contribution to this volume. The calculi Hilbert considers remain very rudimentary until the lectures *1917–18**, where modern mathematical logic is all of a sudden being developed. Back in the Heidelberg talk, the novel syntactic approach to consistency proofs is suggested not for an axiom system concerning real numbers but rather concerning natural numbers; that will be the central topic of the next section. The logical calculus, if it can be called that, is extremely restricted — it is purely equational!

In late 1904, possibly even in early 1905, Hilbert sent a letter to his friend and colleague Adolf Hurwitz. Clearly, he was not content with the proposal he had detailed in Heidelberg:

> It seems that various parties started again to investigate the foundations of arithmetic. It has been my view for a long time that exactly the most important and most interesting questions have not been settled by Dedekind and Cantor (and a fortiori not by Weierstrass and Kronecker). In order to be forced into the position to reflect on these matters systematically, I announced a seminar on the "logical foundations of mathematical thought" for next semester.[43]

[42] *Poincaré 1902*, pp. 252-253

[43] Hilbert mentions in this letter that *Zermelo 1904* had just been published. Zermelos's paper is an excerpt of a letter to Hilbert that was written on 24 September 1904. The German text is from *Dugac*, p. 271: Die Beschäftigung mit den Grundlagen der Arithmetik wird jetzt, wie es scheint, wieder von den verschiedensten Seiten aufgenommen. Dass gerade die wichtigsten u. interessantesten Fragen von Dedekind und Cantor noch nicht (und erst recht nicht von Weierstrass und Kronecker) erledigt worden sind, ist eine Ansicht, die ich schon lange hege und, um einmal in die Notwendigkeit versetzt zu sein, darüber im Zusammenhang nachzudenken, habe ich für nächsten Sommer ein zweistündiges Colleg über die "logischen Grundlagen des math. Denkens" angezeigt.

The lectures from the summer term 1905, *Logische Prinzipien des mathematischen Denkens*, are as special as those of 1904, but for a different reason: one finds in them a critical examination of logical principles and a realization that a broader logical calculus is needed that captures, in particular, universal statements and inferences. Hilbert discusses in some detail (on pp. 253–266) the shortcomings of the approach he pursued in his Heidelberg address; see section 3.1 below. In his subsequent lectures on the foundations of mathematics, Hilbert does not really progress beyond the reflections presented in *1905**. In a letter from Hessenberg to Nelson (dated 7 February 1906 and which I received through Peckhaus), one finds not only a discussion of Russell's early logicism, but strikingly, Dedekind and Hilbert are both described as logicists. Hessenberg remarks:

> Dedekind's standpoint is also a logicist one, but altogether noble and nowhere brazenly dogmatic. One sees and feels everywhere the struggling, serious human being. But given all the superficialities I have found up to now in Russell, I can hardly assume that he is going to provide more than something witty. Zermelo believes that also Hilbert's logicism cannot be carried out, but considers Poincaré's objections as unfounded.[44]

It is only in the Zürich talk *Axiomatisches Denken* that a new perspective emerges. In that pivotal essay, Hilbert remarks that the consistency of the axioms for the real numbers can be reduced, by employing set theoretic concepts, to the very same question for integers; that result is attributed to the theory of irrational numbers developed by Weierstrass und Dedekind. Hilbert continues, insisting on a logicist approach:

> In only two cases is this method of reduction to another more special domain of knowledge clearly not available, namely, when it is a matter of the axioms for the *integers* themselves, and when it is a matter of the foundation of *set theory*; for here there is no other discipline besides logic to which it were possible to appeal.
>
> But since the examination of consistency is a task that cannot be avoided, it appears necessary to axiomatize logic itself and to prove that number theory as well as set theory are only parts of logic.[45]

[44] The German text is: Dedekinds Standpunkt ist ja auch logizistisch, aber durchaus vornehm and nirgends unverfroren dogmatisch. Man sieht und fühlt überall den kämpfenden ernsten Menschen. Aber bei den Oberflächlichkeiten, die ich bis jetzt in Russell gefunden habe, kann ich kaum annehmen, daß er mehr als geistreiches bringt. Zermelo ist der Ansicht, daß auch Hilberts Logizismus undurchführbar ist, hält aber Poincarés Einwendungen für unbegründet.

[45] *Ewald*, p. 1113. Here is the German text from *Hilbert 1918*, p. 153: Nur in zwei Fällen nämlich, wenn es sich um die Axiome der *ganzen Zahlen* selbst und wenn es sich um die Begründung der *Mengenlehre* handelt, ist dieser Weg der Zurückführung auf ein anderes spezielleres Wissensgebiet offenbar nicht gangbar, weil es außer der Logik überhaupt keine Disziplin mehr gibt, auf die alsdann eine Berufung möglich wäre.

Da aber die Prüfung der Widerspruchslosigkeit eine unabweisbare Aufgabe ist, so scheint es nötig, die Logik selbst zu axiomatisieren und nachzuweisen, daß Zahlentheorie sowie Mengenlehre nur Teile der Logik sind.

Hilbert remarks that Russell and Frege provided the basis for this approach. Given Hilbert's closeness to Dedekind's perspective, it is not surprising that he was attracted by the logicist considerations of *Principia Mathematica*. In 1920, however, Hilbert and Bernays reject the logicist program of the Russellian kind. The reason is quite straightforward: for the development of mathematics the axiom of reducibility is needed, and that is in essence a return to existential axiomatics with all its associated problems. In particular, no reduction of epistemological interest has been achieved.

I want to end this section by pointing out that Zermelo's axiomatization of set theory in 1908, inspired by both Dedekind and Hilbert, is another instance of existential axiomatics. Points 1) through 4) in the first section of *Zermelo 1908* are structured in the same way as the axiomatic formulations in Hilbert's *1899* and *1900*. According to Ebbinghaus,[46] Zermelo had intended to give a consistency proof for his axiom system, but was encouraged by Hilbert to publish the axiomatization first and leave the consistency proof for another occasion. Clearly, as to the nature of Zermelo's considerations at that point I am not sure at all. Zermelo just remarks in his paper, "I have not yet even been able to prove rigorously that my axioms are consistent, though this is certainly very essential; instead I have had to confine myself to pointing out now and then that the antinomies discovered so far vanish one and all if the principles here proposed are adopted as a basis."[47] His later investigations on the foundations of set theory in *Zermelo 1930* are parallel to Dedekind's metamathematical reflections described above.

3 DIRECT CONSISTENCY PROOFS

It seems then that Hilbert modified his basic attitude as a *Dedekindian logicist* only after the discovery of Zermelo and Russell's elementary contradiction. That paradox — which had a "catastrophic effect on the mathematical world when it became known" — convinced him that there was a *deep* problem, that difficulties appeared already at an earlier stage, and that the issue had to be addressed in different ways.[48] The notes for the lectures *Zahlbegriff und Quadratur des Kreises* from the summer term of 1904 contain, as we saw already, a detailed discussion of the paradoxes. They examine also Dedekind's contradictory system of all things,

[46] Ebbinghaus quotes in his book 2007, p. 78, from a letter of Zermelo to Hilbert, dated 25 March 1907 in which Zermelo writes: "According to your wish I will finish my set theory as soon as possible, although actually I wanted to include a proof of consistency."

[47] All the remarks from *Zermelo 1908* are found in *van Heijenoort*, pp. 200-201.

[48] In *Über das Unendliche*, p. 169, Hilbert writes, retrospectively: Insbesondere war es ein von Zermelo und Russell gefundener Widerspruch, dessen Bekanntwerden in der mathematischen Welt geradezu von katastrophaler Wirkung war.

That this paradox played a special role is also confirmed by Bernays, who reported to Constance Reid: ... Under the influence of the discovery of the antinomies in set theory, Hilbert temporarily thought that Kronecker had probably be right there. But soon he changed his mind. Now it became his goal, one might say, to do battle with Kronecker with his own weapons of finiteness by means of a modified conception of mathematics. (*Reid*, p. 173)

but both praise his epochal achievements and pointedly insist that fundamental difficulties remain.

> He [Dedekind] arrived at the view that the standpoint of considering the integers as obvious [selbstverständlich] cannot be sustained; he recognized that the difficulties Kronecker saw in the definition of irrationals arise already for integers; furthermore, if they are removed here, they disappear there. This work [*Was sind und was sollen die Zahlen?*] was epochal, but it did not yet provide something definitive, certain difficulties remain. These difficulties are connected, as for the definition of the irrationals, above all to the concept of the infinite; ...[49]

This set the stage for Hilbert's Heidelberg talk in August of that year. Hilbert indicates there a novel way of solving the consistency problem for arithmetic which is taken now in a much more restricted sense, namely, as dealing with just the natural numbers.

3.1 Heidelberg 1904

The shift from arithmetic as pertaining not to the real but to the natural numbers is crucial for Hilbert's presentation in Heidelberg. The consistency of the system considered there would guarantee, as Hilbert puts it, "the consistent existence of the so-called smallest infinite." Its axioms are those for identity extended by Dedekind's requirements for a *simply infinite system*; the induction principle (which follows from Dedekind's minimality condition) is included, but neither formulated properly nor treated in the consistency proof. Using modern notation the axioms that are actually investigated can be given in this way:

1. $x = x$
2. $x = y \;\&\; W(x) \rightarrow W(y)$
3. $x' = y' \rightarrow x = y$
4. $x' \neq 1$

$W(x)$ stands for an arbitrary proposition containing the variable x. The rules, implicit in Hilbert's description of "consequence," are modus ponens and a substitution rule that allows replacing variables by arbitrary sign combinations; other "modes of logical inferences" are only alluded to. Finally, Hilbert most strongly

[49] Here is the German text from *Hilbert 1904**, p. 166: Er drang zu der Ansicht durch, dass der Standpunkt mit der Selbstverständlichkeit der ganzen Zahlen nicht aufrecht zu erhalten ist; er erkannte, dass die Schwierigkeiten, die Kronecker bei der Definition der irrationalen Zahlen sah, schon bei den ganzen Zahlen auftreten und dass, wenn sie hier beseitigt sind, sie auch dort wegfallen. Diese Arbeit war epochemachend, aber sie lieferte doch nichts definitives, es bleiben gewisse Schwierigkeiten übrig. Diese bestehen hier, wie bei der Definition der irrationalen Zahlen, vor allem im Begriff des Unendlichen; ...

emphasized that the problem for the reals is solved once matters are settled for the natural numbers:

> The existence of the totality of real numbers can be demonstrated in a way similar to that in which the existence of the smallest infinite can be proved; in fact, the axioms for real numbers as I have set them up ... can be expressed by precisely such formulas as the axioms hitherto assumed. ... and the axioms for the totality of real numbers do not differ qualitatively in any respect from, say, the axioms necessary for the definition of the integers. In the recognition of this fact lies, I believe, the real refutation of the conception of arithmetic associated with L. Kronecker ...[50]

In the consistency proof for the above system Hilbert uses the property of homogeneity; an equation $a = b$ is called *homogeneous* if and only if a and b have the same number of symbol occurrences. It is easily seen, by induction on derivations, that all equations derivable from axioms (1)–(3) are homogeneous. A contradiction can be obtained only by establishing an unnegated instance of (4) from (1)–(3), but such an instance is necessarily inhomogeneous and thus not provable. Hilbert comments:

> The considerations just sketched constitute the first case in which a direct proof of consistency has been successfully carried out for axioms, whereas the method of a suitable specialization, or of the construction of examples, which is otherwise customary for such proofs - in geometry in particular - necessarily fails here.[51]

Hilbert stressed in his Heidelberg talk the programmatic goal of developing logic and mathematics simultaneously. The actual work has significant shortcomings, as there is no calculus for sentential logic, there is no proper treatment of quantification, and induction is not incorporated into the argument. In sum, there *is* an important shift from semantic arguments to syntactic ones, but the very set-up is inadequate as a formal framework for mathematics. It should also be noted very clearly that the ultimate goal of the consistency proof remains to guarantee the existence of a set, here of the smallest infinite one.

[50] In *van Heijenoort*, pp. 137-8. Here is the German text from *Hilbert 1905*, pp. 278-279: Ähnlich wie die Existenz des kleinsten Unendlich bewiesen werden kann, folgt die Existenz des Inbegriffs der reellen Zahlen: in der Tat sind die Axiome, wie ich sie für die reellen Zahlen aufgestellt habe, genau durch solche Formeln ausdrückbar, wie die bisher aufgestellten Axiome. ... und die Axiome für den Inbegriff der reellen Zahlen unterscheiden sich qualitativ in keiner Hinsicht etwa von den zur Definition der ganzen Zahlen notwendigen Axiomen. In der Erkenntnis dieser Tatsache liegt, wie ich meine, die sachliche Widerlegung der von L. Kronecker vertretenen ... Auffasung der Grundlagen der Arithmetik.

[51] In *van Heijenoort*, p. 135; the German text *loc. cit.* p. 273 is: Die eben skizzierte Betrachtung bildet den ersten Fall, in dem es gelingt, den direkten Nachweis für die Widerspruchslosigkeit von Axiomen zu führen, während die sonst — insbesondere in der Geometrie — für solche Nachweise übliche Methode der geeigneten Spezialisierung oder Bildung von Beispielen hier notwendig versagt.

Hilbert was obviously not fully satisfied with the work he had presented in Heidelberg, and in the summer term of 1905 he did give lectures under the title *Logische Principien des mathematischen Denkens*. Towards the end of the term he reviewed the consistency proof sketched above and considers it "as the first *attempt* [my emphasis, WS] of a consistency proof without giving a special example." He articulates what he considers to be problematic and continues:

> Carrying out this thought in a precise way would at any rate present considerable difficulties, in particular the exact justification of the statement that "all" inferences from the first axioms lead only to homogeneous equations. One arrives with such considerations finally at the thought that proofs should be viewed as mathematical things; the logical calculus indicates that.[52]

But how is it possible to characterize "proofs" mathematically, if they constitute an infinite totality? And, even if one has succeeded to do that, how can one give arguments concerning them? Those questions were already on Hilbert's mind during the summer of 1904; in Born's notes from that term one finds remarks concerning concepts that have particular defining characteristics. Such concepts are viewed as consistent just in case it is not possible to obtain a contradiction by drawing all possible inferences from their characteristics. Hilbert then concludes.

> But in proceeding in this way one obviously uses already arithmetic; in a sense one does so when forming the plural, but certainly when speaking of all inferences etc.[53]

These problematic aspects are sharpened, but not resolved in the 1905 lectures.

3.2 Obstacles

When Hilbert suggests, as I reported above, that proofs might be viewed as "mathematical things", he considers one possibility of doing so "that is more formal and is completely based on the logical calculus." (p. 266) As this formal approach requires a complete formulation of all axioms, in particular, of those that concern "Every", he encounters even there the very same issue. After all, replacing variables by arbitrary combinations is the principal operation and, on account of the great variety of potential combinations, it is difficult to provide a rigorous formulation. Hilbert concludes these reflections by stating:

[52] *Hilbert 1905**, pp. 264-265. The German text is: Eine genaue Durchführung dieses Gedankens würde immerhin noch beträchtliche Schwierigkeiten bieten, insbesondere die exakte Begründung der Aussage, daß "alle" Schlüsse aus den ersten Axiomen nur auf homogene Gleichungen führen. Man kommt bei solchen Überlegungen schließlich darauf, den Beweis selbst als ein mathematisches Ding anzusehen, worauf ja auch der Logikkalkül hinweist.

[53] *Hilbert 1904**, p. 170. The German text is: Dabei benutzt man aber offenbar schon die Arithmetik, eigentlich schon bei der Bildung des Plurals, noch mehr aber, wenn man von allen Schlüssen etc. spricht.

> Accordingly the main task would be to formulate exactly the axioms for "Every" and the concept of combination that is closely connected with them.[54]

As is well known, Poincaré challenged the presumed foundational import of Hilbert's consistency proof, because it proceeds by induction; this is absolutely to the point. He also brought out other *methodological shortcomings* that are closely linked with the attempts Hilbert made in the Heidelberg talk to overcome the difficulties I just sketched.[55]

How penetrating Poincaré's considerations were becomes plain when one reads the remarks of Bernays on Hilbert's Heidelberg paper that give essentially a précis of Poincaré's critique, but also of Hilbert's reflection on the issues; Bernays writes in his 1935:

> In addition, the systematic standpoint of Hilbert's proof theory is not yet fully and clearly developed in the Heidelberg address. Some places indicate that Hilbert wants to avoid the intuitive conception of number and replace it by its axiomatic introduction. Such a procedure would lead to a circle in the proof theoretic considerations.[56]

Bernays emphasizes that Hilbert did not express distinctions that would be central to the later Hilbert Program; for example, Hilbert did not bring out the restriction on the contentual application of the forms of existential and universal judgments.[57]

Poincaré's incisive analysis of the syntactic, proof theoretic approach and his own insight into its shortcomings shifted Hilbert's attention from the stand he had advocated in the Heidelberg talk. Hilbert's lectures, given throughout the period from 1905 to 1917, document his continued foundational concerns. Under the impact of a detailed study of *Principia Mathematica* that began in 1913, Hilbert flirted again with logicism.[58] What resulted from this study, most importantly, were the 1917–18 lectures *Prinzipien der Mathematik* signaling the emergence

[54] *Hilbert 1905**, 269. The German text is: Hiernach wäre dann die exakte Formulierung der Axiome des 'Jedes' und des damit in naher Verbindung stehenden Begriffes der Kombination die Hauptaufgabe.

[55] *Poincaré 1905*, in *Ewald*, pp. 1026-1027.

[56] *Bernays 1935*, p. 200. The German text is: Außerdem ist auch der methodische Standpunkt der Hilbertschen Beweistheorie in dem Heidelberger Vortrag noch nicht zur vollen Deutlichkeit entwickelt. Einige Stellen deuten darauf hin, daß Hilbert die anschauliche Zahlvorstellung vermeiden und durch die axiomatische Einführung des Zahlbegriffs ersetzen will. Ein solches Verfahren würde in den beweistheoretischen Überlegungen einen Zirkel ergeben.

[57] *L.c.*, p. 200. This is part of the following text: Auch wird der Gesichtspunkt der Beschränkung in der inhaltlichen Anwendung der Formen des existentialen und des allgemeinen Urteils noch nicht ausdrücklich und restlos zur Geltung gebracht.

[58] Around 1913 Hilbert became familiar with some of Russell's writings and intended to invite Russell to give lectures in Göttingen; the outbreak of World War I made that impossible. In the notes from the winter term 1914-15 one finds brief remarks about type theory and in serendipitously preserved notes of a student (found in the Institute for Advanced Study in Princeton) even more extended ones. Of greatest significance in indicating the degree to which the Hilbert group discussed the details of *Principia Mathematica* is Behmann's dissertation; that is described in *Mancosu 1999*. Not only the details of *Principia Mathematica* were of significance; on the con-

of modern mathematical logic. In the following two years, the programmatic logicism that was underlying those lectures was abandoned for good reasons. A radical constructivism was adopted instead and subsequently also abandoned – for equally good, but obviously different reasons. The finitist consistency program as we think of it was formulated only in lectures given in the winter term 1921–22.[59]

The evolution towards the program started, however, in the summer term of 1920 when Hilbert came back to the proof theoretic approach of his *1905*. The notes from that term contain in section 7 a consistency proof for exactly the same fragment of arithmetic that was investigated in the Heidelberg talk. Its formulation is informed by the investigations of 1917-18: the language is more properly described; the basic vocabulary consists of variables a, b, ..., non-logical constants 1, +, and all numerals; = and $\neq$ are relation symbols, and $\rightarrow$ is the sole logical symbol. The axioms are:

$$
\begin{aligned}
&1 = 1\\
&a = b \rightarrow a + 1 = b + 1\\
&a + 1 = b + 1 \rightarrow a = b\\
&a = b \rightarrow (a = c \rightarrow b = c)\\
&a + 1 \neq 1
\end{aligned}
$$

Modus ponens and substitution, now applicable only to numerals, are the sole inference rules. The combinatorial argument is different from that given in *1905*, mainly by the introduction of the notion "kürzbar":

> ... If one considers a proof with respect to a particular concrete property it has, then it is possible that the removal of some formulas in this proof still leaves us with a proof that has that particular property. In this case we are going to say that the proof is *kürzbar* with respect to the given property.[60]

Hilbert establishes three lemmata. The first claims that a theorem can contain at most two occurrences of $\rightarrow$, the second asserts that no statement of the form $(A \rightarrow B) \rightarrow C$ can be proved, and the third expresses that a formula $a = b$ is provable only if a and b are the same term. On the surface, Hilbert does not use an induction principle in these proofs. He compares, playing on the double meaning of "kürzbar", the structure of the proof to that for the irrationality of

trary, the broad logicist outlook had a real impact on Hilbert. Hilbert's notes for his course on *Set Theory* from the summer term 1917 and his talk *Axiomatisches Denken* (given on September 11, 1917 in Zürich) reveal a logicist direction of his work.

[59] For details see my *Hilbert's Programs: 1917–1922*, but also for the relationship between the 1917–18 lectures and Hilbert & Ackermann's book of 1928.

[60] *Hilbert 1920**, p. 38. The German text is: Betrachtet man einen Beweis in Hinsicht auf eine bestimmte, konkret aufweisbare Eigenschaft, welche er besitzt, so kann es sein, dass, nach Wegstreichung einiger Formeln in diesem Beweise noch immer ein Beweis (...) übrig bleibt, welcher auch noch jene Eigenschaft besitzt. In diesem Fall wollen wir sagen, dass der Beweis sich in bezug auf die betreffende Eigenschaft kürzen lässt.

$\sqrt{2}$. In that classical indirect argument one starts out by assuming there are p and $q, q \neq 0$, such that $\sqrt{2} = p/q$; without loss of generality p and q are then taken to be relatively prime, i.e., the fraction p/q is not "kürzbar".

The presentation of the consistency proof is simplified in *Hilbert 1922*.[61] Though "Kürzbarkeit" is only implicit in the argument, the strategic point of the modified argument is made quite explicit; Hilbert insists that Poincaré has been refuted.

> Poincaré's objection, claiming that the principle of complete induction cannot be proved but by complete induction, has been refuted by my theory.[62]

Is this to be taken in the strong sense that no induction principle is being used? Or does one use perhaps a special procedure that is based on the construction and deconstruction of numerals and that, by its very nature, is different from the full principle? From a mathematical point of view, Hilbert uses the "least number principle" that is equivalent to induction; it is used in an elementary form for just purely existential statements.

3.3 New directions

In the second part of this paper[63] the theory is significantly expanded. The basic system is modified by replacing the axiom $1 = 1$ with $a = a$, but also by including the axioms $1 + (a + 1) = (1 + a) + 1$ and $a \neq 1 \rightarrow a = \delta(a) + 1$, where δ is the predecessor function. A schema for identity

$$a = b \rightarrow (A(a) \rightarrow A(b))$$

[61] This paper was based on talks given in the spring and summer of 1921 in Copenhagen, respectively Hamburg, where Hilbert presented his *new* investigations on the foundations of arithmetic. Reports on the three Copenhagen talks were published in Berlingske Tidende. Hilbert gave these lectures on March 14, 15, and 17. The first talk had the title *Natur und mathematisches Erkennen*; the second and third were entitled *Axiomenlehre und Widerspruchsfreiheit*. As to the Hamburg talks: Heinrich Behnke gives in his 1976 a brief description of the impact of the Hamburg lecture: "Im Sommer 1922 [sic!] trug Hilbert hier seinen viel zitierten ersten Aufsatz zur Neubegründung der Mathematik vor. Es war eine Entgegnung auf die Schriften von Hermann Weyl und L. E. J. Brouwer, die Unruhe unter die Mathematiker gebracht hatten, weil sie das tertium non datur bestritten und damit schon die Existenz der reellen Zahlen in Frage stellten. In den Diskussionen zu Hilberts Vorträgen [sic!] sprachen Ernst Cassirer, damals Professor der Philosophie in Hamburg, Heinrich Scholz aus Kiel, der spätere Begründer der mathematischen Logik in Deutschland, Kowalewski – Königsberg, der für die Philosophie des "Als ob" von Vaihinger warb, und mancher andere. Die Erörterungen waren viel reger und grundsätzlicher, als es sonst nach mathematischen Vorträgen der Fall zu sein pflegt. Man begriff, daß Hilbert sich um die entscheidende Sicherung der Mathematik gegen Angriffe der Intuitionisten wandte." The detailed information concerning the talks in Copenhagen was conveyed to me by V. Hendricks, that concerning the Hamburg talks by A. Behboud.

[62] *Hilbert 1922*, p. 161. The German text is: Sein [Poincarés] Einwand, dieses Prinzip [der vollständigen Induktion] könnte nicht anders als selbst durch vollständige Induktion bewiesen werden, ist durch meine Theorie widerlegt.

[63] The editors of Hilbert's *Gesammelte Abhandlungen* remark in note 2 on page 168 that the consistency arguments stem from an earlier stage of proof theory, and that a new direction of Hilbert's considerations is indicated on the following pages.

is also added. (Upper case Roman letters are viewed as formula variables.) The system includes rules for universal quantifiers; but as negation is available only for identities, existential quantifiers cannot be introduced. The induction principle and certain logical principles are adjoined to the basic theory. The former principle is given in the form

$$(a)(A(a) \to A(a+1)) \to \{A(1) \to (Z(b) \to A(b))\}.$$

The logical axioms are to reflect the usual inference procedures and are divided into two groups; those in the first group concern the conditional $\to$, and those of the second group deal with negation. Here are the axioms of the first group:

$$\begin{array}{l} A \to (B \to A) \\ \{A \to (A \to B)\} \to (A \to B) \\ \{A \to (B \to C)\} \to \{B \to (A \to C)\} \\ (B \to C) \to \{(A \to B) \to (A \to C)\}. \end{array}$$

Negation is available only for equations. It is thus treated in a restricted way:

$$\begin{array}{l} a \neq a \to A, \\ (a = b \to A) \to ((a \neq b \to A) \to A)). \end{array}$$

As to this restricted use of negation, Hilbert remarks that the system is to be kept constructive; thus, the proper metamathematical direction of the program has not yet been taken.[64] Eight arithmetical axioms concerning Z, addition, and subtraction round out the system. Hilbert emphasizes that the axioms ensure that all statements of arithmetic can be obtained in a formal way.

The editors of Hilbert's *Gesammelte Abhandlungen* mention (in note 1, p. 176) that "a schema for the introduction of functions by recursion equations" has to be

[64]Hilbert emphasizes (on p. 173): Wir haben außer dem Zeichen $\to$ und dem Allzeichen kein anderes logisches Zeichen eingeführt und insbesondere für die logische Operation '*nicht*' die Formalisierung vermieden. Dieses Verhalten gegenüber der Negation ist für unsere Beweistheorie charakteristisch: ein formales Äquivalent für die fehlende Negation liegt lediglich in dem Zeichen $\neq$, durch dessen Einführung die Ungleichheit gewissermaßen ebenso positiv ausgedrückt und behandelt wird, wie die Gleichheit, deren Gegenstück sie ist. Inhaltlich kommt die Negation nur im Nachweise der Widerspruchsfreiheit zur Anwendung, und zwar nur, insoweit es unserer Grundeinstellung entspricht. Mit Rücksicht auf diesen Umstand bringt uns, wie ich glaube, unsere Beweistheorie zugleich auch eine erkenntnistheoretische wichtige Einsicht in die Bedeutung und das Wesen der Negation. When discussing that the existential quantifier can be defined in "formal logic" with the universal quantifier and negation, Hilbert points out (on p. 174): Da aber in unserer Beweistheorie die Negation keine direkte Darstellung haben darf, so wird hier die Formalisierung von 'es gibt' dadurch erreicht, daß man individuelle Funktionszeichen mittels einer Art impliziter Definition einführt, indem gewissermaßen das, 'was es gibt', durch eine Funktion wirklich hergestellt wird. Hilbert discusses then as an example the predecessor function. In the 1923-paper Hilbert acknowledges very briefly on p. 179, note 1: In meiner vorhin zitierten Abhandlung [*Hilbert 1922*; WS] hatte ich dieses Zeichen noch vermieden; es hat sich herausgestellt, daß bei der gegenwärtigen, ein wenig veränderten Darstellung meiner Theorie der Gebrauch des Zeichens 'nicht' ohne Gefahr geschehen kann. This is not a minor shift in perspective, but a methodologically dramatic one!

added, if this last goal is to be reached. As to the extent of the claimed consistency result, they assert (in note 2, p. 176) that it holds only if the universal quantifier is excluded and the induction axiom is replaced by the induction rule. Taking into account these modifications, consistency is claimed for a theory that includes primitive recursive arithmetic. There is no indication of a proof, but the result can obviously not be obtained by extending the combinatorial argument I described above.[65]

Hilbert's *1922* was based, as described in a previous note, on talks he had given in the spring and summer of 1921 in Copenhagen, respectively Hamburg. The first Copenhagen talk has been preserved as a manuscript in Hilbert's own hand. It is worthwhile to quote its last paragraph in order to re-emphasize Hilbert's broad vision for mathematics.

> We went rapidly through those chapters of theoretical physics that are currently most important. If we ask, what kind of mathematics is being considered by physicists, then we see that it is *analysis* that serves physicists in its complete *content* and *extension*. Indeed, it does so in two different ways: first it serves to *clarify* and *formulate* their ideas, and second — as an instrument of calculation — it serves to obtain quickly and reliably numerical results, which help to check the correctness of their ideas. Apart from this face seen by physicists, there is a completely different face that is directed towards philosophy; the features of that face deserve no less our interest. That topic will be discussed in my subsequent talks.[66]

Hilbert did not only expound his particular perspective, but argued also against the stand of Brouwer and Weyl. Summarizing his position in the published essay, he contrasts their constructive tendencies with his own, claiming that Weyl has "failed to see the path to the fulfillment of these [constructive, WS] tendencies" and that "only the path taken here in pursuit of axiomatics will do full justice to

[65] After having made the claim that the consistency of that system can be obtained, Hilbert goes on to say (on p. 176): Aber der wesentlichste Schritt bleibt noch zu tun übrig, nämlich der Nachweis der Anwendbarkeit des logischen Prinzips 'tertium non datur' in dem Sinne der Erlaubnis, auch bei unendlich vielen Zahlen, Funktionen oder Funktionenfunktionen schließen zu dürfen, daß eine Aussage entweder für alle diese Zahlen, Funktionen bzw. Funtionenfunktionen gilt oder daß notwendig unter ihnen eine Zahl, Funktion bzw. Funtionenfunktion vorkommt, für die die Aussage nicht gilt. It is then claimed that by introducing appropriate functionals the foundation of classical analysis can be provided and that the consistency of the extended system can be established.

[66] *Hilbert 1921**, pp. 28-9. The German text is: Wir haben so im Fluge, die gegenwärtig wichtigsten Kapitel der theoretischen Physik durcheilt. Fragen wir, welcher Art Mathematik dabei für den Physiker in Betracht kommt, so sehen wir, dass es die *Analysis* ist, [die] in ihrem gesamten *Inhalt und Umfang* dem Physiker dient und zwar auf zweierlei Weisen: erstens zur *Klärung und Formulierung* seiner Ideen und zweitens als Instrument der Rechnung zur raschen und sicheren Gewinnung numerischer Resultate, durch die er die Richtigkeit seiner Ideen prüft. Ausser diesem Antlitz, in das der Physiker schaut, hat die Mathematik aber noch ein ganz anderes Gesicht, das auf die Philosophie gerichtet ist, [und] dessen Züge nicht minder unser Interesse verdienen. Darüber will ich in meinen folgenden Vorträgen handeln.

the constructive tendencies":

> The goal of securing a firm foundation for mathematics is also my goal. I should like to regain for mathematics the old reputation for incontestable truth, which it appears to have lost as the result of the paradoxes of set theory; but I believe that this can be done while fully preserving its accomplishments. The method I follow in pursuit of this goal is none other than the axiomatic method; its essence is as follows.[67]

Having described the nature of the axiomatic method, he points to the task of recognizing the consistency of the arithmetical axioms including at this point axioms for number theory, analysis, and set theory. This task leads naturally to the metamathematical investigation of suitable formalisms, in which parts of mathematics can be carried out and proofs are given. The concepts of proof and provability are thus "relativized" to the underlying axiom system. Hilbert emphasizes,

> This relativism is natural and necessary; it causes no harm, since the axiom system is constantly being extended, and the formal structure, in keeping with our constructive tendency, is becoming more and more complete.[68]

Hilbert's constructivism comes in at this point in yet a different way, namely, through constructing ever more complete formalisms for the development of mathematics.

The terms "Hilbertsche Beweistheorie" und "finite Mathematik" appear for the first time in the 1921-22 Notes. "Hilbertsche Beweistheorie" is of course self-explanatory, but "finite Mathematik" is not, unless one interprets it simply as having been chosen in contrast to "transfinite Mathematik". Hilbert and Bernays give no philosophical explication. Rather, they insist that an intuitive foundation ("anschauliche Begründung") is to be given and develop paradigmatically elementary number theory on such a foundation; their development is a marvelously detailed and thorough presentation of the beginnings of recursive number theory. The opposition between "finite" and "transfinite Mathematik" is founded on that between "finite" and "transfinite Logik", as the step from "finite" to "transfinite Mathematik" is taken when the classical logical equivalences between the quantifiers are taken to be valid even when we are dealing with infinite totalities. But

[67] *Ewald*, p. 1119. Here is the German text from *Hilbert 1922*, p. 160: Das Ziel, die Mathematik sicher zu begründen, ist auch das meinige; ich möchte der Mathematik den alten Ruf der unanfechtbaren Wahrheit, der ihr durch die Paradoxien der Mengenlehre verloren zu gehen scheint, wiederherstellen; aber ich glaube, daß dies bei voller Erhaltung ihres Besitzstandes möglich ist. Die Methode, die ich dazu einschlage, ist keine andere als die axiomatische; ihr Wesen ist dieses.

[68] *Ewald*, p. 1127. Here is the German text from *Hilbert 1922*, p. 169: Dieser Relativismus ist naturgemäß und notwendig; aus ihm entspringt auch keinerlei Schaden, da das Axiomensystem beständig erweitert und der formale Aufbau, unserer konstruktiven Tendenz entsprechend, immer vollständiger wird.

that cannot be done in "finite Mathematik": "Assuming the unrestricted validity of those equivalences amounts to thinking of totalities with infinitely many individuals as completed ones."[69]

Elementary number theory as Hilbert and Bernays develop it should not be understood as all of their "finite Mathematik". On the contrary, they envision a dramatic expansion; an expansion that is needed, on the one hand, to fully develop analysis and set theory and, on the other hand, to recognize "on the basis of finitist logic" why and to what extent "the application of transfinite inferences [in analysis and set theory] always leads to correct results". The strategy that is suggested reflects what Hilbert had done already, to some extent, around the turn of the century, namely, to shift the Kroneckerian constructivity requirements from mathematics itself to metamathematics. We have to expand, so they demand, the domain of objects that are being considered:

> ... i.e., we have to apply our intuitive considerations also to figures that are not number signs. Thus we have good reasons to distance ourselves from the earlier dominant principle according to which each theorem of pure mathematics is ultimately a statement concerning integers. This principle was viewed as expressing a fundamental methodological insight, but it has to be given up as a prejudice.
>
> We have to adhere firmly to one demand, namely, that the figures we take as objects must be completely surveyable and that only discrete determinations are to be considered for them. It is only under these conditions that our claims and considerations have the same reliability and evidence as in intuitive number theory.[70]

So at this point Hilbert and Bernays find, simultaneously, a resolution for two problems, namely, to fix the standpoint on the basis of which consistency proofs are to be carried out ("finite Mathematik") and on the basis of which the formal development of mathematics can be pushed along in a systematic way. This resolution addresses in particular, so it was thought, the problem that had plagued Hilbert in *1905*, but also before and after, namely, to capture the content of a

[69] *Hilbert 1921-22**, p. 2a of Part III. The German text is: Denn die Annahme der unbeschränkten Gültigkeit jener Aequivalenzen kommt darauf hinaus, dass wir uns Gesamtheiten von unendlich vielen Individuen als etwas fertig Vorhandenes denken.

[70] *Hilbert 1921–22**, pp. 4a and 5a of Part III. The German text is: ... d.h. wir müssen unsere anschaulichen Ueberlegungen auch auf andere Figuren als auf Zahlzeichen anwenden. Wir sehen uns somit veranlasst, von dem früher herrschenden Grundsatz abzugehen, wonach jeder Satz der reinen Mathematik letzten Endes in einer Ausage über ganze Zahlen bestehen sollte. Dieses Prinzip, in welchem man eine grundlegende methodische Erkenntnis erblickt hat, müssen wir als Vorurteil preisgeben.

An einer Forderung aber müssen wir festhalten, dass nämlich die Figuren, die wir als Gegenstände nehmen, vollständig überblickbar sind, and dass an ihnen nur *diskrete* Bestimmungen in Betracht kommen. Denn nur unter diesen Bedingungen können unsere Behauptungen und Ueberlegungen die gleiche Sicherheit und Handgreiflichkeit haben wie in der anschaulichen Zahlentheorie.

statement concerning all natural numbers or all proofs of a formal system without needing the "axiomatic definition of the system (Gesamtbegriffes) of natural numbers [or of proofs, WS] as one thing." A systematic formal development had already been required in *1905** (p. 250) but was viewed as extremely difficult. It was hoped then, that it would lead to a complete description of mathematical and thus of all theoretical thinking; in addition, the effectiveness requirements for the metamathematical notions would achieve an almost Leibnizian end:

> ... and in this way, by writing down certain symbols in limited number, every question shall be decidable by an "experiment with chalk or pen".[71]

To recognize early and deep roots of principled aspects of Hilbert's thinking seems to me to be important; it is equally important to see how the study of *Principia Mathematica* opened a new way to the formalization of mathematics that had not been available to Hilbert in *1905*, in spite of Frege's work. What is the status of "finite" in "finite Mathematik" in that historical regard? Does it indicate a new philosophical perspective with a particular foundational claim that emerged in the early 1920s? The way in which the concept is introduced in *1921–22**, very matter of fact and without great fanfare, almost leads one to suspect that Hilbert and Bernays employ a familiar one.[72] That suspicion is hardened by *aspects of the past* and an *attitude* that is pervasive until 1932: as to the attitude, finitism and intuitionism were considered as co-extensional until Gödel and Gentzen proved in 1932 the consistency of classical arithmetic relative to its intuitionistic version; as to aspects of the past, there are Hilbert's many and oft-repeated remarks about constructivist demands, mostly connected to Kronecker. Hilbert himself remarked in 1931 that Kronecker's conception of mathematics "essentially coincides with our finitist mode of thought."[73]

There is also a very informative paper by Felix Bernstein, entitled "Die Mengenlehre Georg Cantors und der Finitismus" and published in 1919; it is one of the few papers listed at the end of the Introduction to the second edition of *Principia Mathematica* in 1927. Bernstein characterizes "Finitismus" as a familiar foundational stance that opposed set theory since its very beginnings in Cantor's and Dedekind's work. Somewhat indiscriminately, Kronecker and Hermite, Borel and Poincaré, Richard and Lindelöf, but also Brouwer and Weyl (on account of *Das Kontinuum*) are viewed as members of the finitist movement. Bernstein's paper was published before Hilbert and Bernays viewed "finite Mathematik" as

[71] *Hilbert 1905**, p. 275. The quotation is part of the following German text: Jeder mathematische Satz und Beweis soll auf eine hinschreibbare Form gebracht werden, und jede Frage soll damit durch ein Hinschreiben gewisser Zeichen in begrenzter Zahl, durch ein "Experiment mit Kreide oder Feder" entschieden werden können.

[72] It is most informative to look at the "Disposition" (outline) for the 1921-22 lectures. Bernays proposed it in a letter to Hilbert dated 17 October 1921. The actual lectures followed that "Disposition". But what is in the lectures called "finitist" is here still simply called "constructive"; the outline is reprinted in full in *Sieg 1999,* pp. 35-36.

[73] *Ewald*, p.1151.

foundationally significant for proof theory. (The characteristic features of finitism Bernstein formulated are actually close to those of "finite Mathematik" Hilbert and Bernays emphasized in *1921–22**.)

That concrete background of the term "finitism" could of course be a topic of detailed historical analysis, but it won't be discussed any further here. I just state as a fact that in the lecture notes from the 1920s no detailed discussion of "finite Mathematik" is found. The most penetrating analysis is given in *Bernays 1930* still emphasizing the co-extensionality of finitism and intuitionism. Indeed, Bernays interprets Brouwer's mathematical work as showing that considerable parts of analysis and set theory can be "given a finitist foundation".[74] What is found in the lecture notes is innovative metamathematical work that tries to overcome in leaps and bounds the obstinate difficulties of giving consistency proofs for stronger formal theories, but in the end unsuccessfully. That's what I am going to describe in some detail now.

4 FINITIST PROOF THEORY

In the winter term 1921-22, Hilbert and Bernays not only introduced the terms "Hilbertsche Beweistheorie" and "finite Mathematik", but they also used novel techniques to give a finitist proof of the restricted consistency result formulated in *Hilbert 1922*. The proof is presented in Kneser's *Mitschrift* of these lectures. Indeed, the proof theoretic considerations start on 2 February and end on 23 February 1922; they include a treatment of definition by (primitive) recursion and proof by induction. The latter principle is formulated as a rule and restricted to quantifier-free formulas. The part of the official notes *1921–22** that present the consistency proof was written by Bernays after the term had ended and contains a different argument; that argument pertains only to the basic system and is sketched in *Hilbert 1923*. I am going to describe a third modification found in Kneser's *Mitschrift* of the 1922-23 lectures. Apart from this important beginning of proof theoretic investigations, I will discuss Hilbert's *Ansatz* for the treatment of quantifiers, some relevant further developments in the 1920s due to Ackermann and von Neumann, and the refined and systematic presentation of this work in the first and second volume of *Grundlagen der Mathematik*.

[74] *Bernays 1930*, pp. 41-42. The quoted remark is part of this longer passage: Brouwer hat den Standpunkt Kroneckers nach zwei Richtungen weitergeführt: einerseits in Hinsicht auf die philosophische Motivierung, durch die Aufstellung seiner Theorie des "Intuitionismus", andererseits dadurch, daß er gezeigt hat, wie man im Gebiete der Analysis und der Mengenlehre den finiten Standpunkt zur Anwendung bringen und, durch eine Umgestaltung der Begriffsbildung und der Schlussweisen von Grund auf, diese Theorien wenigstens zu einem beträchtlichen Teil auf finitem Wege begründen kann.

Bernays attached to "Intuitionismus" in the above passage a footnote, saying: Es scheint mir im Interesse der Klärung der Diskussion angezeigt, den Ausdruck "Intuitionismus" als Bezeichnung einer philosophischen Ansicht zu gebrauchen, im Unterschied von dem Terminus "finit", der eine bestimmte Art des Schließens und der Begriffsbildung bezeichnet.

4.1 Proof transformations

The work in 1921–22 gave a new direction to proof theoretic investigations and required the partial resolution of some hard technical, as well as deep methodological issues. In the first group of issues we find in particular the treatment of quantifiers; that is absolutely crucial, if consistency is to be established for the kind of formalisms that had been used in the 1917–18 lectures to represent classical mathematical practice. The second group of issues concerns the extent of metamathematical means (is contentual induction admitted?) and the character of the formal systems to be investigated (are they to be quasi-constructive?). The last issue has a direct resolution: it is addressed in *1921–22** by a modification of the logical calculus, i.e., by replacing the (in-) equalities in the axioms for negation by (negations of) arbitrary formulas B. The axioms are given in the form:

$$B \to (\neg B \to A)$$
$$(B \to A) \to ((\neg B \to A) \to A)$$

With this emendation we have the system that is used in Hilbert's *1923*; the first axiom expresses the *principle of contradiction* and the second the *principle of tertium non datur*.

Though the technical resolution is straightforward it required a new philosophical perspective that is, perhaps first and best, articulated in Bernays' *1922* that was written during the early fall of 1921. The paper starts out with a discussion of *existential axiomatics*, which presupposes, as we saw in Part 2, a system of objects satisfying the structural conditions formulated by the axioms. The assumption of such a system contains, Bernays writes, "something so-to-speak transcendent for mathematics." Thus, the question arises, "which principled position with respect to it should be taken." Bernays remarks that it might be perfectly coherent to appeal to an intuitive grasp of the natural number sequence or even of the manifold of real numbers. However, that could obviously not be an intuition in any primitive sense, and one should take very seriously the tendency in the exact sciences to use, as far as possible, only restricted means for acquiring knowledge. Bernays is thus led to a programmatic demand:

> Under this perspective we are going to try, whether it is not possible to give a foundation to these transcendent assumptions in such a way that only *primitive intuitive knowledge is used*.[75]

Clearly, contentual mathematics is to be based on primitive intuitive knowledge and that includes for Bernays already at this stage induction.

Bernays' perspective is clearly reflected in the notes *1921–22**. The notes contain a substantial development of finitist arithmetic that involves from the very outset contentual induction ("anschauliche Induktion"). This step was not a small

[75] *Bernays 1922*, p. 11. The German text is: Unter diesem Gesichtspunkt werden wir versuchen, ob es nicht möglich ist, jene transzendenten Annahmen in einer solchen Weise zu begründen, daß nur *primitive anschauliche Erkenntnisse zur Anwendung kommen*.

one for Hilbert. In 1904 he had described Kronecker as a dogmatist because of his view "that integers — and in fact integers as a general notion (parameter value) — are directly and immediately given." In the intervening years he scolded Kronecker for just that reason and accused him of not being radical enough. Bernays reports that Hilbert viewed this expansion of contentual arithmetic as a *compromise*, and it is indeed a compromise between a radical philosophical position and intrinsic demands arising from metamathematical investigations. But it is also a compromise in that it gives up the direct parallelism with the case of physics, when statements that can be verified by observation are viewed as corresponding to statements that can be checked by calculation.

For the central steps of the consistency argument we need a definition: a formula is called *numeric* if it is built up solely from $=, \neq$, numerals, and sentential logical connectives. The first step transforms formal proofs with a numeric endformula into proofs that contain only closed formulas, without changing the endformula. In the second step these modified proofs are turned into configurations that may no longer be proofs, but consist only of numeric formulas; that is achieved by reducing the closed terms to numerals. Finally, all formulas in these configurations are brought into disjunctive normal form and syntactically recognized as "true". Assume now that there is a formal proof of $0 \neq 0$; the last observation allows us to infer that the endformula of this proof is true. However, $0 \neq 0$ is not true and hence not provable.

This proof is given with many details in *1921–22**. It is sketched in *Hilbert 1923* (on p. 184) and presented in a slightly modified form, according to Kneser's *Mitschrift*, in the winter term of 1922–23. For that term the papers *Bernays 1922*, *Hilbert 1922*, and *Hilbert 1923* were all available as indicated explicitly by Kneser. The system in *Kneser 1922-23* uses a calculus for sentential logic that incorporates the connectives & and $\vee$. The main point of this expanded language and calculus is to provide a more convenient formal framework for representing informal arguments. The axioms for the additional connectives are given as follows:

$$
\begin{array}{ll}
A\&B \to A & A\&B \to B \\
A \to (B \to A\&B) & \\
A \to A \vee B & B \to A \vee B \\
(A \to C) \to ((B \to C) \to (A \vee B \to C)) &
\end{array}
$$

Clearly, the axioms correspond directly to "natural" rules for these connectives, and one finds here the origin of Gentzen's natural deduction calculi. In *Bernays 1927b* the completeness of this calculus is asserted and its methodological advantages are discussed. Bernays writes:

> The initial formulas can be chosen in quite different ways. In particular one has tried to get by with the smallest number of axioms, and in this respect the limit of what is possible has indeed been reached. One supports the purpose of logical investigations better, if one separates, as it is done in the axiomatics of geometry, different groups of axioms in

> such a way that each of these groups expresses the role of one particular logical operation.[76]

Then he lists the three groups, namely, axioms for the conditional $\rightarrow$, for conjunction $\&$, disjunction $\vee$, as well as for negation $\neg$. This formal system is used in *Ackermann 1925*, in *Über das Unendliche*, and in Hilbert's second Hamburg talk of 1927. Hilbert gives in the Hamburg talk an interestingly different formulation of the axioms for negation, calling them the *principle of contradiction* and the *principle of double negation*:

$$(A \rightarrow (B\&\neg B)) \rightarrow \neg A$$
$$\neg\neg A \rightarrow A$$

These are axiomatic formulations of Gentzen's natural deduction rules for classical negation; the earlier axioms for negation can be derived easily. (This new formulation can be found in *van Heijenoort* on p. 382 and more fully on pp. 465-466.)

The consistency argument for the system of arithmetic described above, but with this logical calculus, is given in *Kneser 1922-23* (pp. 20-22); its form is used in the first volume of *Grundlagen der Mathematik*. It begins by separating linear proofs into proof threads ("Auflösung in Beweisfäden"). Then variables are eliminated in the resulting proof trees ("Ausschaltung der Variablen"), and the numerical value of functional terms is determined by calculation ("Reduktion der Funktionale"). That makes it possible to determine the truth or falsity of formulas in the very last step. Here that is not done by turning formulas into disjunctive normal form, but rather simply by truth-table computations ("Wahrheitsbewertung"). Incorporating definition by recursion and proof by induction into these considerations is fairly direct.

From a contemporary perspective this argument shows something very important. If a formal theory contains a certain class of finitist functions, then it is necessary to appeal to a wider class of functions in the consistency proof: an *evaluation function* is needed to determine uniformly the numerical value of terms, and such a function is no longer in the given class. The formal system considered in the above consistency proof includes primitive recursive arithmetic, and the consistency proof goes beyond the means available in primitive recursive arithmetic. At this early stage of proof theory, finitist mathematics is consequently stronger than primitive recursive arithmetic. This assessment of the relative strength is sustained, as we will see, throughout the development reported in this essay.

[76] *Bernays 1927b*, p. 10. The German text is: Die Wahl der Ausgangsformeln kann auf sehr verschiedene Weise getroffen werden. Man hat sich besonders darum bemüht, mit einer möglichst geringen Zahl von Axiomen auszukommen, und hat hierin in der Tat die Grenze des Möglichen erreicht. Den Zwecken der logischen Untersuchung wird aber besser gedient, wenn wir, entsprechend wie in der Axiomatik der Geometrie, verschiedene *Axiomengruppen* voneinander sondern, derart, daß jede von ihnen die Rolle einer logischen Operation zum Ausdruck bringt. It is informative to compare these remarks with Hilbert's in the Introduction to *1899*, quoted in section 2.2.

4.2 Hilbert's Ansatz: the ε-substitution method

Ackermann reviewed in section II of his *1925* the above consistency proof; the section is entitled *The consistency proof before the addition of the transfinite axioms* ("Der Widerspruchsfreiheitsbeweis vor Hinzunahme der transfiniten Axiome"). The very title reveals the restricted programmatic significance of this result, as it concerns a theory that is part of finitist mathematics and need not be secured by a consistency proof. The truly expanding step that goes beyond the finitist methods just described involves the treatment of quantifiers. Already in Hilbert's *1922* we find a brief indication of his *Ansatz*, which is elaborated in *Kneser 1921–22*. Indeed, the treatment of quantifiers is just the first of three steps that have to be taken: the second concerns the general induction principle for number theory, and the third an expansion of proof theoretic considerations to analysis.

Hilbert sketched in his Leipzig talk of September 1922, which was published as *Hilbert 1923*, the consistency proof from section 4.1. However, the really dramatic aspect of his proof theoretic discussion is the treatment of quantifiers with the τ-function, the dual of the later ε-operator. The logical function τ associates with every predicate $A(a)$ a particular object $\tau_a(A(a))$ or simply τA. It satisfies the *transfinite axiom* $A(\tau A) \rightarrow A(a)$, which expresses, according to Hilbert, "if a predicate A holds for the object τA, then it holds for all objects a." The τ-operator allows the definition of the quantifiers:

$$(a)A(a) \leftrightarrow A(\tau A)$$
$$(Ea)A(a) \leftrightarrow A(\tau(\neg A))$$

Hilbert extends the consistency argument then to the "first and simplest case" that goes beyond the finitist system: this "Ansatz" will evolve into the ε-substitution method. It is only in the Leipzig talk that Hilbert gave proof theory its principled formulation and discussed its technical tools — for the first time outside of Göttingen. As far as analysis is concerned, Hilbert had taken as the formal frame for its development ramified type theory with the axiom of reducibility in *1917–18**. In his Leipzig talk he considers a third-order formulation; appropriate functionals (Funktionenfunktionen) allow him to prove (on pp. 189-191) i) the least upper bound principle for sequences and sets of real numbers and ii) Zermelo's choice principle for sets of sets of real numbers. He conjectures that the consistency of the additional transfinite axioms can be patterned after that for τ. He ends the paper with the remark:

> Now the task remains of precisely carrying out the basic ideas I just sketched; its solution completes the founding of analysis and prepares the ground for the founding of set theory.[77]

In Supplement IV of the second volume of *Grundlagen der Mathematik*, Hilbert and Bernays give a beautiful exposition of analysis and emphasize, correctly, that

[77] *Hilbert 1923*, p. 191. The German text is: Es bleibt nun noch die Aufgabe einer genauen Ausführung der soeben skizzierten Grundgedanken; mit ihrer Lösung wird die Begründung der Analysis vollendet und die der Mengenlehre angebahnt sein

the formalisms for their "deductive treatment of analysis" were used in Hilbert's early lectures on proof theory and that they were described in *Ackermann 1925*.

The above concrete proof theoretic work was directly continued in Ackermann's thesis, on which *Ackermann 1925* is based. The paper was submitted to Mathematische Annalen on 30 March 1924 and published in early 1925.[78] It starts out, as I mentioned, with a concise review of the earlier considerations for quantifier-free systems. The ε-symbol had been substituted for the τ-symbol in 1923 as is clear from *Kneser 1922-23*, and in Ackermann's dissertation the ε-calculus as we know it replaced the τ-calculus from 1922. The transfinite axiom of the ε-calculus is obviously dual to the axiom for τ and is formulated in *Ackermann 1925* in section III entitled, "Die Widerspruchsfreiheit bei Hinzunahme der transfiniten Axiome und der höheren Funktionstypen." For number variables the crucial axiom is formulated as $A(a) \to A(\varepsilon A)$ and allows of course the definition of quantifiers:

$$(Ea)A(a) \leftrightarrow A(\varepsilon A)$$
$$(a)A(a) \leftrightarrow A(\varepsilon(\neg A))$$

The remaining transfinite axioms are adopted from *Hilbert 1923*. However, the ε-symbol is actually characterized as the least-number operator, and the recursion schema with just number variables is extended to a schema that permits also function variables.

The connection to the mathematical development in Hilbert's paper is finally established in section IV, where Ackermann explores the "Tragweite" of the axioms. At first it was believed that Ackermann had completed the task indicated by Hilbert at the end of the Leipzig talk and had established not only the consistency of full classical number theory but even of analysis. However, a note was added in proof (p. 9) that significantly restricted the result; von Neumann noted a mistake in Ackermann's arguments. In his own paper *Zur Hilbertschen Beweistheorie*, submitted on 29 July 1925, a consistency proof was given for a system that covers, von Neumann asserts (p. 46), Russell's mathematics without the axiom of reducibility or Weyl's system in *Das Kontinuum*. Ackermann corrected his arguments and obtained an equivalent result; that corrected proof is discussed in *Hilbert 1927* and *Bernays 1927a*.[79] (It is also the basis for the corresponding presentation in the second volume of *Hilbert & Bernays*; see section 4.4 below.) In his Bologna talk of 1928 Hilbert stated, in line with von Neumann's observation, that the consistency of full number theory had been secured by the proofs of Ackermann and von Neumann; according to Bernays in his preface to the second volume of *Grundlagen der Mathematik* that belief was sustained until 1930.

A number of precise metamathematical problems were formulated in Hilbert's Bologna talk, among them the completeness question for first-order logic, which

[78] In his lecture *Über das Unendliche* given on 4 June 1925 in Münster, Hilbert presented an optimistic summary of the state of proof theoretic affairs together with bold new considerations. Even before this talk, Hilbert had turned his attention to issues beyond the consistency of number theory and analysis, namely, the solution of the continuum problem.

[79] See *van Heijenoort*, p. 477 and p. 489, but also note 3 on p. 489 and the introductions to these papers.

is presented in a new formulation suggested by the Löwenheim Skolem Theorem. The latter theorem also suggested, so it seems, that higher mathematical theories might not be syntactically complete. In any event, Hilbert's Bologna address shows most dramatically through its broad perspective and clear formulation of open problems, how far mathematical logic had been moved in roughly a decade. This is a remarkable achievement with impact not just on proof theory; the (in-) completeness theorems, after all, are to be seen in this broader foundational enterprise.[80]

In terms of specifically proof theoretic results no striking progress was made, in spite of the fact that most talented people had been working on the consistency program. That state of affairs had intrinsic reasons. In the preface to the first volume of *Grundlagen der Mathematik*, Bernays asserts that a presentation of proof theoretic work had almost been completed in 1931. But, he continues, the publication of papers by Herbrand and Gödel produced a deeply changed situation that resulted in an extension of the scope of the work and its division into two volumes. The eight chapters of the first volume can be partitioned into three groups: chapters 6 to 8 give refined versions of results that had been obtained in the 1920s, investigating the consistency problem and other metamathematical questions for a variety of (sub) systems of number theory; chapters 3 to 5 develop systematically the logical framework of first-order logic (with identity); chapters 1 and 2 discuss the central foundational issues and the finitist standpoint in most informative and enlightening ways. Let me start with a discussion of the central issues as viewed in 1934.

4.3 Grundlagen, 1934

Chapter 1 begins with a general discussion of axiomatics, at the center of which is a distinction between *contentual* and *formal axiomatic theories*. Contentual axiomatic theories like Euclid's geometry, Newton's mechanics, and Clausius' thermodynamics draw on experience for the introduction of their fundamental concepts and principles, all of which are understood contentually. Formal axiomatic theories like Hilbert's axiomatization of geometry, by contrast, abstract away such intuitive content. They begin with the assumption of a fixed system of things (or several such systems), which is delimited from the outset and constitutes a "domain of individuals for all predicates from which the statements of the theory are built up." The existence of such a domain of individuals constitutes an "idealizing assumption that joins the assumptions formulated in the axioms."[81] Hilbert and

[80] A remark indicating the depth of Dedekind's continued influence: Hilbert formulated as Problem I of his Bologna talk the consistency of the ε-axioms for function variables and commented, "The solution of problem I justifies also Dedekind's ingenious considerations in his essay *Was sind und was sollen die Zahlen?*"

[81] *Hilbert and Bernays 1934*, pp. 1-2. The last quotation is part of this German text: In der Voraussetzung einer solchen Totalität des "Individuen-Bereiches" liegt — abgesehen von den trivialen Fällen, in denen eine Theorie es ohnehin nur mit einer endlichen, festbegrenzten Gesamtheit von Dingen zu tun hat — eine idealisierende Annahme, die zu den durch die Axiome

Bernays refer to this approach as *existential axiomatics*. They clearly consider formal axiomatics to be a sharpening of contentual axiomatics, but are also quite explicit that the two types of axiomatics complement each other and are both necessary.

The consistency of a formal axiomatic theory with a finite domain can be established by exhibiting a model satisfying the axioms. However, for theories F with infinite domains consistency proofs present a special problem, because "reference to non-mathematical objects cannot settle the question whether an infinite manifold exists; the question must be solved within mathematics itself."[82] One must treat the consistency problem for such a theory F as a *logical problem* — from a proof theoretic perspective. This involves formalizing the principles of logical reasoning and proving that from F one cannot derive (with the logical principles) both a formula and its negation. Proof theory need not prove the consistency of every F; it suffices to find a proof for a system F with a structure that is (i) surveyable to make a consistency proof for the system plausible and (ii) rich enough so that the existence of a model S for F guarantees the satisfiability of axiom systems for branches of physics and geometry. Arithmetic (including number theory and analysis) is considered to be such an F.

For a consistency argument to be foundationally significant it must avoid the idealizing existence assumptions made by formal axiomatic theories. But if a proof theoretic justification of arithmetic by elementary means were possible, might it not also be possible to develop arithmetic directly, free from non-elementary assumptions and thus not requiring any additional foundational justification? The answer to this question involves elementary presentations of parts of number theory and formal algebra; these presentations serve at the same time to illuminate the finitist standpoint. Finitist deliberations take here their purest form, i.e., the form of "*thought experiments* involving objects assumed to be concretely given."[83] The word finitist is intended to convey the idea that a consideration, a claim or definition respects (i) that *objects are representable*, in principle, and (ii) that *processes are executable*, in principle.[84]

A direct, elementary justification for all of mathematics is not possible, Hilbert and Bernays argue, because already in number theory and analysis one uses non-finitist principles. While one might circumvent the use of such principles in number

formulierten Annahmen hinzutritt.

[82] L.c., p. 17. The German text is: Auf Grund dieser Überlegungen kommen wir zu der Einsicht, daß die Frage nach der Existenz einer unendlichen Mannigfaltigkeit durch eine Berufung auf außermathematische Objekte nicht entschieden werden kann, sondern innerhalb der Mathematik selbst gelöst werden muß.

[83] L.c., p. 20. The German text is: Das Kennzeichnende für diesen methodischen Standpunkt ist, daß die Überlegungen in der Form von *Gedankenexperimenten* an Gegenständen angestellt werden, die als *konkret vorliegend* angenommen werden.

[84] L.c., p. 32. The German text is: Im gleichen Sinne wollen wir von finiten Begriffsbildungen und Behauptungen sprechen, indem wir allemal mit dem Wort "finit" zum Ausdruck bringen, daß die betreffende Überlegung, Behauptung oder Definition sich an die Grenzen der grundsätzlichen Vorstellbarkeit von Objekten sowie der grundsätzlichen Ausführbarkeit von Prozessen hält und sich somit im Rahmen konkreter Betrachtung vollzieht.

theory (as only the existence of the domain of integers is assumed), the case for analysis is different. There the objects of the domain are themselves infinite sets of integers, and the principle of the excluded middle is applied in this extended domain. Thus one is led back to using finitist proof theory as a tool to secure the consistency of mathematics. (This restriction is relaxed only at the end of the second volume when "extensions of the methodological framework of proof theory" are considered.) The first stage of this project, the proof theoretic investigation of appropriate logical formalisms, occupies chapters 3 through 5. These systems are close to contemporary ones; they can be traced back to the lectures given in 1917–18 and were presented also in *Hilbert and Ackermann 1928*. The second stage is carried out in chapters 6 through 8 and involves the investigation of sub-systems of first-order number theory; there are three different groups of such systems.

The first group consists of weak fragments of number-theory that contain few, if any, function symbols and extend predicate logic with equality by axioms for 0, successor and $<$; some of the fragments involve quantifier-free induction. Meta-mathematical relations between them are explored, and independence as well as consistency results are established. The main technique for giving consistency proofs is that discussed in section 4.1 above. However, since the formalisms contain quantifiers, an additional procedure is required, namely a procedure that assigns *reducts*, quantifier-free formulas acting as witnesses, to formulas containing quantifiers. The method underlying this procedure is due to Herbrand and Presburger. Consistency is inferred from results that involve the notion of *verifiability*, extending the notion of truth to formulas containing free variables, bound variables, and recursively defined function signs.[85] In order to establish the consistency of a formalism F, one proves that every formula not containing formula variables is verifiable, if it is derivable in F. Since $0 \neq 0$ is not verifiable, it is not derivable in F; thus, F is consistent.

The second group of subsystems of number theory contains formalisms arising from the elementary calculus with free variables (the quantifier free fragment of the predicate calculus) through the addition of functions defined by recursion. Hilbert and Bernays discuss at the beginning of chapter 7 the formalization of the definition principle for recursion, reminding the reader how they viewed recursive definition of a function in their discussion of finitist number theory in chapter 1. There they emphasize that by "function" they understand an intuitive instruction ("anschauliche Anweisung") on account of which a numeral is associated with a given numeral, or a pair of numerals, or a triple ... of numerals. The proof method of complete induction is not considered as an "independent principle" but

[85]More precisely, letting A be a formula of the formalism F: (i) if A is a numeric formula, it is verifiable if it is true; (ii) if A contains free numeric variables (but no formula variables or bound variables), it is verifiable if one can show by finitist means that the substitution of arbitrary numerals for variables (followed by the evaluation of all function-expressions and their replacement through their numerical values) yields a true numeric formula; (iii) if A contains bound variables but no formula variables, it is verifiable if its reduct is verifiable (according to (i) and (ii)).

as a consequence that is gathered from the concrete build-up of numerals.[86] Similarly, definition by recursion is not viewed as an independent definition principle but rather as "an agreement concerning an abbreviated description of certain construction processes by means of which one obtains a numeral from one or several given numerals."[87] These processes or procedures can be mimicked formally by the introduction of function symbols together with recursion equations.[88] They treat *in a first step* ("zunächst") what they consider as the simplest schema of recursion (and "restrict *for the time being* the notion of recursion to it"). That simplest schema of recursion is the schema of what we (and they later on p. 326/331) call *primitive recursion*, namely,

$$f(a,\ldots,k,0) = \mathfrak{a}(a,\ldots,k),$$
$$f(a,\ldots,k,n') = \mathfrak{b}(a,\ldots,k,n,f(a,\ldots,k,n)).$$

$\mathfrak{a}$ and $\mathfrak{b}$ denote previously defined functions and $a,\ldots,k,n$ are numerical variables. What then is the formalization of the intuitive procedure associated with a recursive definition that is formally reflected by the recursion schema for f? It is given by a derivation — via the recursion equations and the axioms for identity — of the equation $f(a,\ldots,k,z) = t(a,\ldots,k)$ for each numeral z, where t is a term not containing f. Then they argue that the value of $f(a,\ldots,k,z)$ can be determined for each sequence of numerals $a,\ldots,k$ and conclude their considerations with: "Thus we can completely mimick the recursive calculation procedure of finitist number theory in our formalism by the deductive application of the recursion equations."[89]

They then prove on pp. 298–300 a general *Consistency Theorem*: *Let F be a formalism extending the elementary calculus with free variables by verifiable axioms (that may contain recursively defined functions whose defining equations are taken as axioms) and the schema of quantifier-free induction, then every derivable formula of F (without formula variables) is verifiable.* This theorem is explicitly taken to establish the consistency of a number of formalisms including in particular that of recursive number theory. Having formally developed the latter theory to indicate the strength of this recursive treatment of number theory, they state:

[86] *Hilbert and Bernays 1934*, p. 25. The German text is: Wir haben es also hier nicht mit einem selbständigen Prinzip zu tun, sondern mit einer Folgerung, die wir aus dem konkreten Aufbau der Ziffern entnehmen.

[87] *Hilbert and Bernays 1934*, p. 27. The German text is: Es handelt sich hier bei der Definition durch Rekursion wiederum nicht um ein selbständiges Definitionsprinzip, sondern die Rekursion hat im Rahmen der elementaren Zahlentheorie lediglich die Bedeutung einer Vereinbarung über eine abgekürzte Beschreibung gewisser Bildungsprozesse, durch die man aus einer oder mehreren gegebenen Ziffern wieder eine Ziffer enthält.

[88] *Hilbert and Bernays 1934*, p. 287. This is a translation of part of the following German text: Dieses Verfahren [der Berechnung von Funktionswerten in der finiten Zahlentheorie] können wir auch im Formalismus nachbilden, indem wir allgemein die Einführung von Funktionszeichen in Verbindung mit Rekursionsgleichungen zulassen.

[89] *Hilbert and Bernays 1934*, p. 292. The German text is: Somit können wir das rekursive Berechnungsverfahren der finiten Zahlentheorie in unserem Formalismus durch die deduktive Anwendung der Rekursionsgleichungen vollkommen nachbilden.

> This recursive number theory is closely related to intuitive number theory, as we have considered it in §2, because all of its formulas can be given *a finitist contentual interpretation.* This contentual interpretability follows from the verifiability of all derivable formulas of recursive number theory, as we established it already. Indeed, in this area verifiability has the character of a direct contentual interpretation, and thus the proof of consistency could be obtained here so easily.

As we have a finitist consistency proof for recursive arithmetic, finitist mathematics here goes beyond recursive arithmetic. Their notion of recursive number theory was at first, as I emphasized above, to involve only primitive recursions; so finitist mathematics is stronger than primitive recursive arithmetic. Indeed, after the above remark they continue the discussion of the relationship between intuitive and recursive number theory as follows:

> Recursive number theory is differentiated from intuitive number theory by its formal restrictions; its only method for forming concepts, apart from explicit definitions, is the recursion schema, and its methods of derivation are also firmly delimited.
>
> However, without taking away from recursive number theory what is the characteristic feature of its method, we can admit certain *extensions of the recursion schema* as well as of the induction schema. We'll discuss these [extensions] now briefly.

The extending formalisms involve the schema of nested recursion ("verschränkte Rekursion") that allows in particular the definition of the Ackermann function; they remark that their previous consistency results are extended to these proper extensions of recursive number theory as well; thus, the bounds of finitist mathematics are pushed still further.[90]

The third group of formalisms that are actually equivalent to full number theory is investigated towards the end of chapter 7 and in chapter 8. The first of these

[90]The two long quotations above are both found on p. 325/330 of *Hilbert and Bernays 1934*. The remark concerning the finitist provability of the consistency of the extended formalisms is on p. 325/330; more precisely, it is asserted there, all formulas (not containing formula variables) that are provable in the extended formalisms are verifiable. Here is the first German text:

Diese rekursive Zahlentheorie steht insofern der anschaulichen Zahlentheorie, wie wir sie im §2 betrachtet haben, nahe, als ihre Formeln sämtlich *einer finiten inhaltlichen Deutung fähig* sind. Diese inhaltliche Deutbarkeit ergibt sich aus der bereits festgestelletn Verifizierbarkeit aller ableitbaren Formeln der rekursiven Zahlentheorie. In der Tat hat in diesem Gebiet die Verifizierbarkeit den Charakter einer direkten inhaltlichen Interpretation, und der Nachweis der Widerspruchsfreiheit war daher auch hier so leicht zu erbringen.

And here is the second passage:

Der Unterschied der rekursiven Zahlentheorie gegenüber der anschaulichen Zahlentheorie besteht in ihrer formalen Gebundenheit; sie hat als einzige Methode der Begriffsbildung, außer der expliziten Definition, das Rekursionsschema zur Verfügung, und auch die Methoden der Ableitung sind fest umgrenzt.

Allerdings können wir, ohne der rekursiven Zahlentheorie das Charakteristische ihrer Methode zu nehmen, gewisse *Erweiterungen des Schemas der Rekursion* sowie auch des Induktionsschemas zulassen. Auf diese wollen wir noch kurz zu sprechen kommen.

is the formalism of the axiom system (Z); call this formalism Z. Hilbert and Bernays comment that the techniques used in the consistency proofs for fragments of number theory cannot be generalized to Z, as any reduction procedure for Z would provide a decision procedure for Z and, thus, allow solving all number theoretic problems. They leave the possibility of such a procedure as an open problem (whose solution, if it exists, they view as a long way off) and focus on examining whether Z provides the means for formalizing informal number theory. With this end in mind, it is proved in chapter 8 that all recursive functions are representable in Z. For this proof they establish three separate claims: (1) the least number operator μ can be explicitly defined in terms of Russell and Whitehead's ι-symbol; (2) any recursive definition can be explicitly defined in Z_μ, i.e., Z extended by defining axioms for the μ-operator; (3) the addition of the ι-rule to Z is a conservative extension of Z.[91] Volume I concludes with the remark that the above results entail the consistency of Z_μ relative to that of Z, but that none of the results or methods considered so far suffice to show that Z is consistent.

4.4 Limited results for quantifiers

The consistency proofs in chapter 7.a) of the first volume are given for quantifier-free systems. In the second volume, these theories are embedded in the system of full predicate logic together with the ε-axioms in the form $A(a) \rightarrow A(\varepsilon_x.A(x))$. The first crucial task is to eliminate all references to bound variables from proofs of theorems that do not contain them; axioms used in these proofs must not contain bound variables either. In the formulation of Hilbert and Bernays, the consistency of a system of proper axioms relative to the predicate calculus together with the ε-axioms is to be reduced to the consistency of the system relative to the elementary calculus (with free variables).[92] The consistency of the latter system is recognized on account of a suitable finitist interpretation. Thus, it is emphasized, operating with the ε-symbol can be viewed as "merely an auxiliary calculus, which is of considerable advantage for many metamathematical considerations."[93]

In the framework of the extended calculus bound variables are associated with just ε-terms, as the quantifiers can be defined. The initial elimination result is

[91] The second claim is first established for primitive recursively defined functions, but then extended to the case of nested recursion; cf. *Hilbert and Bernays 1934*, p. 421/431. The latter result is attributed to von Neumann and Gödel. As to the general issues compare von Neumann's letter to Gödel of 29 November 1930; the letter is reprinted and translated in volume V of Gödel's *Collected Works*.

[92] *Hilbert and Bernays 1939*, p. 33. Referring to the first ε-theorem they write there: Die Bedeutung dieses Theorems besteht ja darin, daß es die Frage der Widerspruchsfreiheit eines Systems von eigentlichen Axiomen ohne gebundene Variablen bei Zugrundelegung des Prädikatenkalkuls und des ε-axioms zurückführt auf die seiner Widerspruchsfreiheit bei Zugrundelegung des elementaren Kalkuls mit freien Variablen.

[93] *Hilbert and Bernays 1939*, p. 13. The quotation is part of this German text: Es besteht ja keineswegs die Notwendigkeit, das ε-symbol in den endgültigen deduktiven Aufbau des logisch-mathematischen Formalismus einzubeziehen. Vielmehr kann das Operieren mit dem ε-Symbol als ein bloßer Hilfskalkul angesehen werden, der für viele metamathematische Überlegungen von erheblichem Vorteil ist.

the *First ε-Theorem*: *If the axioms $A_1, \ldots, A_k$ and the conclusion of a proof do not contain bound individual variables or (free) formula variables, then all bound variables can be eliminated from the proof.* The argument can be extended to cover proofs of purely existential formulas, but the proofs yield then as their conclusion suitable disjunctions of instances of the endformula. Based on this extension a *Consistency Theorem* of a quite general character is proved: *If the axioms $A_1, \ldots, A_k$ are verifiable, then (i) any provable formula containing at most free individual variables is verifiable, and (ii) for any provable, purely existential formula $(Ex_1)\ldots(Ex_n)A(x_1, \ldots, x_n)$* (with only the variables shown) there are variable-free terms $t_1, \ldots, t_n$ *such that $A(t_1, \ldots, t_n)$* is true. This theorem is applied to establish the consistency (i) of Euclidean and Non-Euclidean geometry without continuity assumptions in section 1.4, and (ii) of arithmetic with recursive definitions, but only quantifier-free induction in sections 2.1 and 2.2. In essence then, the consistency theorem from *Herbrand 1931* (cf. section 5.1) has been reestablished in a subtly more general way: Hilbert and Bernays allow the introduction of a larger class of recursive functions.[94] Putting the result in the appropriate systematic context, we see that the consistency proof of *1922* for the quantifier-free system of primitive recursive arithmetic has been extended to cover that system's expansion by full classical quantification theory. Indeed, the extensions from *1934* are also covered.

The remainder of chapter 2 discusses the difficulty of extending the elimination procedure in the proof of the first ε-theorem to a system with full induction and examines Hilbert's original *Ansatz* for eliminating ε-symbols.[95] The next two chapters investigate the formalism for predicate logic and begin in chapter 3 with a proof of the *Second* ε-Theorem: *If the axioms and the conclusion of a proof (in predicate logic with identity) do not contain ε-symbols, then all ε-symbols can be eliminated from the proof.* Herbrand's theorem is obtained as well as a variety of criteria for the refutability of formulas in predicate logic; proofs of the Löwenheim–Skolem theorem and of Gödel's completeness theorem are also given. These considerations are used to establish results concerning the decision problem; solvable cases as well as reduction classes are discussed. In chapter 4 Gödel's "arithmetization of metamathematics" is presented in great detail and applied to obtain a fully formalized proof of the completeness theorem.

The completeness theorem can be taken as stating that the consistency of an axiom system relative to the calculus of predicate logic coincides with satisfiability of the system by an arithmetic model. The formalized proof is intended to

[94] *Hilbert and Bernays 1939*, p. 52. That is made explicit in note 1 on page 52. Though this larger class of functions (all general recursive ones?) allows a generalized notion of verifiability, the latter would no longer guarantee a finitist consistency proof. They write there: Diese [allgemeine Anweisung zur Einführung von Funktionszeichen in *Herbrand 1931*] ist insofern etwas enger als die hier formulierte Anweisung, als Herbrand verlangt, daß das Verfahren der Wertbestimmung sich, so wie bei den Schematen der rekursiven Definition, durch eine finite Deutung der Axiome ergeben soll.

[95] As to the character of the original and the later version of the elimination method and Ackermann's work see pp. 21, 29-30, 92ff, 121ff, the note on p. 121, as well as Bernays' preface.

establish a kind of finitist equivalent to a consequence of this formulation, namely, that the consistency relative to the predicate calculus guarantees consistency, as it is formulated on p. 205, in an open contentual sense ("im unbegrenzten inhaltlichen Sinne"). The finitist equivalent is formulated roughly as follows: if a formula is irrefutable in predicate logic, then it remains irrefutable in "every consistent number theoretic formalism," i.e., in every formalism that is consistent and remains consistent when the axioms of Z_μ and verifiable formulas are added. That result expresses a strong deductive closure of the predicate calculus, but only if Z_μ is consistent.[96] Thus, there is another reason for establishing finitistically the consistency of this number theoretic formalism.

5 INCOMPLETENESS

The second volume picks up where the first left off. As we saw, it presents in chapters 1 and 2 Hilbert's proof theoretic "Ansatz" based on the ε-symbol as well as related consistency proofs; this is the first main topic. The methods used there open a simple approach to Herbrand's theorem, which is at the center of chapter 3. The discussion of the decision problem at the end of that chapter leads, after a thorough discussion of the "method of the arithmetization of metamathematics," in the next chapter to a proof theoretic sharpening of Gödel's completeness theorem. The remainder of the second volume of *Grundlagen der Mathematik* is devoted to the second main topic, the examination of the fact, "which is the basis for the necessity to expand the frame of the contentual inference methods, which are admitted for proof theory, beyond the earlier delimitation of the 'finitist standpoint'."[97] Of course, Gödel's incompleteness theorems are at the center of that discussion, and they were also at the center for the changed situation that forced a rethinking of the proof theoretic enterprise. This rethinking took two directions, (i) exploring the extent of finitist mathematics, and (ii) demarcating the appropriate methodological standpoint for proof theory.

5.1 Changed situation

The consistency result in *Herbrand 1931* took an additional step beyond Ackermann's, as far as the very claim is concerned; at the time, it was the strongest proof theoretic result established by finitist means. Comparing Ackermann's result with his own, Herbrand wrote to Bernays in a letter dated 7 April 1931:

[96] *Hilbert and Bernays 1939*, p. 253/263. Here is the German text of which the quoted remark is a part: Das hiermit bewiesene Theorem hat allerdings seine Bedeutung als ein Vollständigkeitssatz, d.h. als Ausdruck einer Art von deduktiver Abgeschlossenheit des Prädikatenkalkuls, nur unter der Voraussetzung der Widerspruchsfreiheit des zahlentheoretischen Formalismus.

[97] Bernays wrote in "Zur Einführung" of the second volume of *Grundlagen der Mathematik*: Das zweite Hauptthema bildet die Auseinandersetzung des Sachverhalts, auf Grund dessen sich die Notwendigkeit ergeben hat, den Rahmen der für die Beweistheorie zugelassenen inhaltlichen Schlußweisen gegenüber der vorherigen Abgrenzung des "finiten Standpunktes" zu erweitern.

> In my arithmetic the axiom of complete induction is restricted, but one may use a variety of functions other than those that are defined by simple recursion: in this direction, it seems to me, that my theorem goes a little farther than yours [i.e., Ackermann's].

Gödel formulated Herbrand's central result in this beautiful and penetrating way: "If we take a theory which is constructive in the sense that each existence assertion made in the axioms is covered by a construction, and if we add to this theory the non-constructive notion of existence and all the logical rules concerning it, e.g., the law of excluded middle, we shall never get into any contradiction."[98] Obviously, that formulation covers equally well the above result obtained by Hilbert and Bernays in *1939*. When Herbrand wrote his last paper *1931*, he knew already of Gödel's incompleteness theorems and von Neumann's related conjecture. The latter had drawn the consequences of Gödel's results most sharply and dramatically: "If there is a finitist consistency proof at all, then it can be formalized. Therefore, Gödel's proof implies the impossibility of any [such] consistency proof."[99]

Was this the end of the proof theoretic program? An answer to this question required an answer to another question: What is finitist mathematics by means of which the proof theoretic program was supposed to be carried out? There is, as I described above, no sharp and clear answer to be found in Hilbert and Bernays' work from the 1920s as to the upper limits. At the very start of Hilbert's new proof theoretic investigations, finitist considerations are extended beyond primitive recursive arithmetic, and it is clear that finitist arguments make use only of proof by induction and definition by recursion within a quantifier-free framework. That is clear also from the perspective of "outsiders" in the early 1930s, cf. *Herbrand 1931* and *Gödel 1933*. At issue was, and to some extent still is, what are the recursion schemata that are finitistically allowed? Again, there is no sharp and clear answer to be extracted (or to be expected) from the early work of Hilbert and Bernays. However, one fact can be stated: von Neumann, Herbrand, Gödel in *1933* and even Hilbert and Bernays in *1934* and *1939* consider the Ackermann function as a finitist one. I.e., the lower bound on the strength of finitist methods is given by a proper extension of primitive recursive arithmetic. The issue of an upper bound was much discussed in 1931 and subsequent years, when people tried to assess the impact of the incompleteness theorems. I mentioned already von Neumann's immediate and decidedly pessimistic judgment.

Gödel was much more cautious in his *1931* and also in his letter to Herbrand dated 25 July 1931; in the former he asserted that the second incompleteness theorem does not contradict Hilbert's formalist viewpoint:

---

[98] *Gödel 1933*, p. 52.

[99] Minutes of the meeting of the Vienna Circle on 15 January 1931; the minutes are found in the Carnap Archives of the University of Pittsburgh. The German text is: Wenn es einen finiten Widerspruchsbeweis überhaupt gibt, dann lässt er sich auch formalisieren. Also involviert der Gödelsche Beweis die Unmöglichkeit eines Widerspruchsbeweises überhaupt. The minutes are quoted more fully in *Sieg 1988*, note 11, and in *Mancosu 1999a*, pp. 36-37.

> ... this viewpoint presupposes only the existence of a consistency proof in which nothing but finitist means of proof are used, and it is conceivable that there exist finitist proofs that cannot be expressed in the formalism of P (or of M or A).[100]

Both von Neumann and Herbrand argued in correspondence with Gödel against that position and conjectured that finitist arguments may be formalizable already in elementary number theory (and if not there, then undoubtedly in analysis). That would establish the limited reach of Hilbert's finitist consistency program.[101] What is the perspective Hilbert and Bernays took on these deliberations in volume II of *Grundlagen der Mathematik*?

5.2 Gödel's theorems

Hilbert and Bernays' discussion of the incompleteness theorems begins with a thorough investigation of semantic paradoxes. This investigation does not try to "solve" the paradoxes in the case of natural languages, but focuses on the question under what conditions analogous situations can occur in the case of *formalized languages*. These conditions are formulated quasi-axiomatically for deductive formalisms F taking for granted that there is a bijection between the expressions of F and natural numbers, a "Gödel-numbering." The formalism F and the numbering are required to satisfy roughly two *representability conditions*: $R1$) primitive recursive arithmetic is "contained in" F, and $R2$) the syntactic properties and relations of F's expressions, as well as the processes that can be carried out on them, are given by primitive recursive predicates and functions.

For the consideration of the first incompleteness theorem the second representability condition is made more specific. It is now required that the *substitution function* (yielding the number of the expression obtained from an expression with number **k**, when every occurrence of the number variable a is replaced by a numeral **l**) is given primitive recursively by a binary function s(k,l) and the *proof predicate* by a binary relation $B(m, n)$ (holding when **m** is the number of a sequence of formulas that constitutes an F-derivation of the formula with number **n**). Consider, as Gödel did, the formula $\neg B(m, s(a, a))$; according to the first representability condition this is a formula of the formalism F and has a number, say **p**. Because of the defining property of $s(k, l)$, the value of s(**p**,**p**) is then the number **q**

[100] *Gödel 1931*, p. 194. The German text is: Es sei ausdrücklich bemerkt, daß Satz XI (und die entsprechenden Resultate über M, A) in keinem Widerspruch zum Hilbertschen formalistischen Standpunkt stehen. Denn dieser setzt nur die Existenz eines mit finiten Mitteln geführten Widerspruchsfreiheitsbeweises voraus und es wäre denkbar, daß es finite Beweise gibt, die sich in P (bzw. M, A) *nicht* darstellen lassen. The formalism P is that of Principia Mathematica, M is that of von Neumann's set theory, and A that of analysis. The letter to Herbrand is found in volume V of Gödel's *Collected Works*.

[101] That discussion in their correspondence with Gödel is described in *Sieg 2005*. There is also an informative letter of Herbrand to his friend Claude Chevalley; the letter was written on 3 December 1930, when Herbrand was in Berlin working on proof theory and discussing Gödel's theorems with von Neumann. (The letter is presented and analyzed in the Appendix of *Sieg 1994*.)

of the formula $\neg B(m, s(\mathbf{p}, \mathbf{p}))$. The equation $s(\mathbf{p}, \mathbf{p}) = \mathbf{q}$ is provable in F; thus, $\neg B(m, s(\mathbf{p}, \mathbf{p}))$ is actually equivalent to $\neg B(m, \mathbf{q})$ and expresses that "the formula with number $\mathbf{q}$ is not provable in F." As $\mathbf{q}$ is the number of $\neg B(m, s(\mathbf{p}, \mathbf{p}))$, this formula consequently expresses (via the equivalence) its own underivability. The argument adapted from that for the liar paradox leads, assuming that this formula is provable, directly to a contradiction in F. But instead of encountering a paradox one infers that the formula is not provable, if the formalism F is consistent.

Hilbert and Bernays discuss — following Gödel and assuming the ω-consistency of F — the unprovability of the sentence $\neg(x)\neg B(x, \mathbf{q})$. Then they establish the Rosser version of the first incompleteness theorem, i.e., the independence of a formula R from F assuming just F's consistency. Thus, a "sharpened version" of the theorem can be formulated for deductive formalisms satisfying certain conditions: *One can always determine a unary primitive recursive function f, such that the equation $f(m) = 0$ is not provable in F, while for each numeral $\boldsymbol{l}$ the equation $f(\boldsymbol{l}) = 0$ is true and provable in F; neither the formula $(x)f(x) = 0$ nor its negation is provable in F.*[102] This sharpened version of the theorem asserts that every sufficiently expressive, sharply delimited, and consistent formalism is deductively incomplete.

For a formalism F that is consistent and satisfies the restrictive conditions, the proof of the first incompleteness theorem shows the formula $\neg B(m, \mathbf{q})$ to be unprovable. However, it also shows that the sentence $\neg B(\mathbf{m}, \mathbf{q})$ holds and is provable in F, for each numeral $\mathbf{m}$. The second incompleteness theorem is obtained by formalizing these considerations, i.e. by proving in F the formula $\neg B(m, \mathbf{q})$ from the formal expression C of F's consistency. That is possible, however, only if F satisfies additional conditions, the so-called *derivability conditions*. The formalized argument makes use of the representability conditions R1) and R2). The second condition now requires also that there is a unary primitive recursive function e, which when applied to the number $\mathbf{n}$ of a formula yields as its value the number of the negation of that formula. The derivability conditions are formulated as follows: D1) If there is a derivation of a formula with number $\mathbf{l}$ from a formula with number $\mathbf{k}$, then the formula $(Ex)B(x, \mathbf{k}) \rightarrow (Ex)B(x, \mathbf{l})$ is provable in F; D2) The formula $(Ex)B(x, e(k)) \rightarrow (Ex)B(x, e(s(k, l)))$ is provable in F; D3) If $f(m)$ is a primitive recursive term with m as its only variable and if $\mathbf{r}$ is the number of the equation $f(a) = 0$, then the formula $f(m) = 0 \rightarrow (Ex)B(x, s(\mathbf{r}, m))$ is provable in F. Consistency is formally expressed by $(Ex)B(x, n) \rightarrow \neg(Ex)B(x, e(n))$; starting with that formula C as an assumption, the formula $\neg B(m, \mathbf{q})$ is obtained in F by a rather direct argument. So, in case the formalism F is consistent, no formalized proof of consistency, i.e., no derivation of the formula C can exist in F.[103]

There are two brief remarks with which I want to complement this metamathematical discussion of the incompleteness theorems. The first simply states that verifying the representability conditions and the derivability conditions is the central mathematical work that has to be done; Hilbert and Bernays accomplish this

[102] *Hilbert and Bernays 1939*, p.279/288.
[103] *Hilbert and Bernays 1939,* pp. 286-288/296-297.

for the formalisms Z_μ and Z.[104] Thus, the second volume of *Grundlagen der Mathematik* contains the first full argument for the second incompleteness theorem; after all, Gödel's paper contains only a minimal sketch of a proof. However, it has to be added — and that is the second remark — that the considerations are not fully satisfactory for a *general* formulation of the theorems, as there is no argument given why deductive formalisms should satisfy the particular restrictive conditions on their syntax. (This added observation points to one of the general methodological issues discussed briefly at the end of section 5.3.)

Existential formal axiomatics emerged in the second half of the 19^{th} century and found its remarkable expression in 1899 through Hilbert's *Grundlagen der Geometrie*; its existential assumption constituted the really pressing issue for the various Hilbert programs during the period from 1899 to 1934, the date of the publication of the first volume of *Grundlagen der Mathematik*. The finitist consistency program began to be pursued in 1922 and is the intellectual thread holding the investigations in both volumes together. The ultimate goal of proof theoretic investigations, as Hilbert formulated it in the preface to volume I, is to recognize the usual methods of mathematics, without exception, as consistent. Hilbert continued,

> With respect to this goal I would like to emphasize the following: the view, which temporarily arose and maintained that certain recent results of Gödel imply the infeasibility of my program, has been shown to be erroneous.[105]

How is the program affected by those results? Is it indeed the case, as Hilbert expressed it also in that preface of *1934*, that Gödel's theorems just force proof theorists to exploit the finitist standpoint in a sharper way?

5.3 Completeness and bounds on finitism

The second question is raised prima facie through the second incompleteness theorem. However, Hilbert and Bernays discuss also the effect of the first incompleteness theorem and ask explicitly, whether the deductive completeness of formalisms is a necessary feature for the consistency program to make sense. They touched on this very issue already in pre-Gödel publications: Hilbert in his Bologna talk of 1928, and Bernays in his penetrating article *1930*. Hilbert formulated in his talk the question of the syntactic completeness for number theory and analysis as Problem IV; he concluded the discussion by suggesting that "in höheren Gebieten" (higher than number theory) it is thinkable that a system of axioms could be consistently extended by a statement S, but also by its negation $\neg S$; the acceptance of

[104] The considerations for the former systems start on p. 293/306, for the latter on p. 324/337.

[105] *Hilbert and Bernays 1934*, in Hilbert's "Zur Einführung". The German text is: Im Hinblick auf dieses Ziel möchte ich hervorheben, daß die zeitweilig aufgekommene Meinung, aus gewissen neueren Ergebnissen von Gödel folge die Undurchführbarkeit meiner Beweistheorie, als irrtümlich erwiesen ist.

one of the statements is then to be justified by "systematic advantages (principle of the permanence of laws, possibilities of further developments etc.)."[106]

Hilbert conjectured that number theory is deductively complete. That is reiterated in Bernays' *1930* and followed by the remark that "the problem of a real proof for this is completely unresolved." The problem becomes even more difficult, Bernays continues, when we consider systems for analysis or set theory. However, this "Problematik" is not to be taken as an objection against the standpoint presented:

> We only have to realize that the [syntactic] formalism of statements and proofs we use to represent our conceptions does not coincide with the [mathematical] formalism of the structure we intend in our thinking. The [syntactic] formalism suffices to formulate our ideas of infinite manifolds and to draw the logical consequences from them, but in general it [the syntactic formalism] cannot combinatorially generate the manifold as it were out of itself.[107]

That is also the central point in the general discussion of the first incompleteness theorem. Indeed, Hilbert and Bernays emphasize that in formulating the problems and goals of proof theory they avoided from the beginning "to introduce the idea of a total system for mathematics with a philosophically principled significance." It suffices to characterize the actual systematic structure of analysis and set theory in such a way that it provides an appropriate frame for (the reducibility of) the geometric and physical disciplines.[108] That point was already emphasized in section 3.3.

From these reflective remarks it follows that the first incompleteness theorem for the central formalisms F (of number theory, analysis, and set theory) does not directly undermine Hilbert's program. It raises nevertheless in its sharpened form a peculiar issue: any finitist consistency proof for F would yield a finitist proof

[106] *Hilbert 1930*, p. 6. The German text is: In höheren Gebieten wäre der Fall der Widerspruchsfreiheit von S und der von $\neg S$ denkbar; alsdann ist die Annahme einer der beiden Aussagen S, $\neg S$ als Axiom durch systematische Vorzüge (Prinzip der Permanenz von Gesetzen, weitere Aufbaumöglichkeiten u.s.w.) zu rechtfertigen. This paragraph does not appear in the original proceedings of the Bologna Congress in 1928.

[107] *Bernays 1930*, p. 59. The German text is: Wir müssen uns nur gegenwärtig halten, daß der Formalismus der Sätze und Beweise, mit denen wir unsere Ideenbildung zur Darstellung bringen, nicht zusammenfällt mit dem Formalismus derjenigen Struktur, die wir in der Gedankenbildung intendieren. Der Formalismus reicht aus, um unsere Ideen von unendlichen Mannigfaltigkeiten zu formulieren und aus diesen die logischen Konsequenzen zu ziehen, aber er vermag im allgemeinen nicht, die Mannigfaltigkeit gleichsam aus sich kombinatorisch zu erzeugen.

[108] *Hilbert and Bernays 1939*, p. 280/289. The extended German text is: Wir haben in unserer Darstellung der Ausgangsproblematik und der Zielsetzung der Beweistheorie von vornherein vermieden, den Gedanken eines Totalsystems der Mathematik in einer philosophisch prinzipiellen Bedeutung einzuführen, vielmehr uns begnügt, die tatsächlich vorhandene Systematik der Analysis und Mengenlehre als eine solche zu charakterisieren, die einen geeigneten Rahmen für die Einordnung der geometrischen und physikalischen Disziplinen bildet. Diesem Zweck kann ein Formalismus auch entsprechen, ohne die Eigenschaft der vollen deduktiven Abgeschlossenheit zu besitzen.

of a statement in recursive number theory — that is not provable in F. Finitist methods would thus go beyond those of analysis and set theory, even for the proof of number theoretic statements. (That is of course the situation Gödel contemplated in his remarks, when claiming that his results don't contradict Hilbert's standpoint; see section 5.1.) This is a "paradoxical" situation, in particular, as Hilbert and Bernays quite unambiguously state in the first volume (on p. 42), "finitist methods are included in the usual arithmetic." Consequently, even the first theorem forces us to address two general tasks, namely, (i) exploring the extent of finitist methods, and (ii) demarcating appropriately the methodological standpoint for proof theory.

Tasks (i) and (ii) are usually associated with the second incompleteness theorem, which, as emphasized in section 5.2, allows us to infer directly and sharply that a finitist consistency proof for a formalism F (satisfying the representability and derivability conditions) cannot be carried out in F. Hilbert and Bernays explore the extent of finitist methods in chapter 5.a) by first trying to answer the question, in which formalism their various finitist investigations can actually be carried out. The immediate claim is that most considerations can be formalized, perhaps with a great deal of effort, already in primitive recursive arithmetic. But then they assert: "At various places this formalism is admittedly no longer sufficient for the desired formalization. However, in each of these cases the formalization is possible in (Z_μ)."[109] They point to the more general recursion principles from chapter 7 of the first volume as an example of "procedures of finitist mathematics" that cannot be captured in primitive recursive arithmetic, but can be formalized in Z_μ.

In the remainder of chapter 5.3.a) they discuss "certain other typical cases," in which the boundaries of primitive recursive arithmetic are too narrow to allow a formalization of their prior finitist investigations. There is, first of all, the issue of an evaluation function that is needed for the consistency proof of primitive recursive arithmetic (already in volume I), but cannot be defined by primitive recursion (p. 341/355). Secondly, there is the general concept of a calculable function (p. 341/356). That concept is used (p. 189/198) to formulate a sharpened notion of satisfiability, i.e. *effective satisfiability*, in their treatment of solvable cases of the decision problem. Thirdly, they discuss (p. 344/358) the principle of induction for universally quantified formulas used in consistency proofs. The issue surrounding this principle is settled metamathematically, as we know, by later proof theoretic work: the system of elementary number theory with this induction principle is conservative over recursive arithmetic, whether in the narrow or wider

[109] *Hilbert and Bernays 1939*, p. 340/354. The extended German text is: An verschiedenen Stellen ist freilich dieser Formalismus [der rekursiven Zahlentheorie mit nur primitiver Rekursion, WS] nicht mehr für die gewünschte Formalisierung ausreichend. Doch zeigt sich dann jedesmal die Möglichkeit der Formalisierung in (Z_μ). Gewisse über die rekursive Zahlentheorie (im ursprünglichen Sinne) hinausgehende Verfahren der finiten Mathematik haben wir bereits im §7 besprochen, nämlich die Einführung von Funktionen durch verschränkte Rekursionen und die allgemeineren Induktionsschemata. Dabei erwähnten wir auch die Formalisierbarkeit dieser Rekursions- und Induktionsschemata im vollen zahlentheoretischen Formalismus.

sense.[110]

As to the second issue (effective satisfiability), some remarks concerning Supplement II are relevant here, because the notion of a calculable function has to be sharpened in such a way as to be formalizable. The presentation of the negative solution of the decision problem in supplement II is preceded by an analysis of the concept "reckonable function", i.e., of a function whose values can be calculated according to elementary rules. The latter rather vague notion is sharpened in a way that is methodologically very similar to their analysis of the incompleteness theorems, namely, by formulating *recursiveness conditions* for deductive formalisms that allow equational reasoning. The central condition requires the proof predicate to be primitive recursive. It is then shown that the functions calculable in formalisms satisfying these conditions are exactly the general recursive ones. Though their analysis is not fully satisfactory for the reason mentioned in section 5.2, it is nevertheless a major and concluding step for the analysis of effectively calculable functions as pursued in the mid-1930s by Gödel, Church, Kleene, and others. It should be emphasized that Hilbert and Bernays are perfectly clear about one fact, namely, that this class of functions goes beyond the class of functions introduced in Herbrand's 1931 by the axioms of his "Groupe C". Indeed, they write:

> General instructions for the introduction of new function symbols were given by Herbrand in his paper [*1931*, WS] ... ("Groupe C"). These are somewhat narrower than the ones formulated here inasmuch as Herbrand requires that the procedure for determining values [of functions] is to follow from a finitist interpretation of the axioms, as is the case for the schemata of recursive definitions.[111]

So it is also here quite clear that Hilbert and Bernays view Herbrand's procedure of introducing function symbols as a definitely finitist one; recall that Herbrand mentions explicitly that a symbol for the Ackermann function can be introduced in this way.

5.4 Methodological frame

The careful re-examination of their own proof theoretic practice leads Hilbert and Bernays to the conclusion that some finitist considerations go beyond primitive recursive arithmetic, but can be formally captured in Z_μ; most of this was pointed out already in volume I. It is at exactly this point that the second incompleteness theorem provides, as the title of Chapter 5 states, the "reason for extending the methodological frame for proof theory." Already in the transition from Chapter

[110] See *Sieg 1991*, section 2.1 and references there to work by Charles Parsons.

[111] Here is the German text: Eine allgemeine Anweisung zur Einführung neuer Funktionszeichen durch Axiome wurde von Herbrand in seiner Abhandlung [*1931*, WS]... gegeben ("Groupe C"). Diese ist insofern etwas enger als die hier formulierte Anweisung, als Herbrand verlangt, daß das Verfahren der Wertbestimmung sich, so wie bei den Schematen der rekursiven Definition, durch eine finite Deutung der Axiome ergeben soll. This is found in note 2 on p. 52 (of the second volume).

4 to Chapter 5, Hilbert and Bernays state specifically that consequences of the theorem force us to view the domain of the contentual inference methods used for the investigations of proof theory more broadly "than it corresponds to our development of the finitist standpoint so far."[112]

The question is, whether there are any methods that can still be called properly "finitist" and yet go beyond Z_μ. Hilbert and Bernays argue that this is not a precise question, as "finitist" is not a sharply delimited notion, but rather indicates methodological guidelines that enable us to recognize some considerations as definitely finitist and others as definitely non-finitist. The limits of finitist considerations are to be "loosened"; two possibilities for such loosenings are considered and quickly seen to be conservative.[113] Which further loosening is "admissible, if we want to adhere to the fundamental tendencies of proof theory?" Against this background two results, then quite recent, are examined: the reduction of classical arithmetic Z to the system $\mathbf{Z}$ of arithmetic with just minimal logic, and Gentzen's consistency proof for a version of $\mathbf{Z}$ (and thus of Z) using a special form of transfinite induction.

The reductive result Hilbert and Bernays formulate is a slightly stronger one than the one obtained by Gödel and, independently, by Gentzen. The proof showing that Z is consistent relative to $\mathbf{Z}$ is an elementary finitist one. Thus, the obstacle for obtaining a finitist consistency proof for Z does not lie in the fact that it contains the typically non-finitist logical principles like tertium non datur! The obstacle appears already when one tries to give a finitist consistency proof for $\mathbf{Z}$. The consistency of Z would be established on the basis of any assumptions, "which suffice to give a verifying interpretation of the restricted formalism."[114] Such a contentual verification, based on interpretations of Kolmogoroff and Heyting, is then examined with the conclusion that it involves the intuitionistic understanding of negation as absurdity. In using the underlying contentual concept of consequence, it is claimed, "we are totally turning away from Hilbert's methodological ideas for proof theory."[115] That is consonant with the view expressed in the first volume

[112] *Hilbert and Bernays 1939*, p. 253/263. The German text is: Dieser Umstand [daß die bisherigen Methoden nicht zum Nachweis der Konsistenz des vollen zahlentheoretischen Formalismus ausreichen, WS] ... findet nun eine grundsätzliche Erklärung durch ein Theorem von Gödel über deduktive Formalismen, für welches der zahlentheoretische Formalismus einen ersten Anwendungsfall bildet und dessen Konsequenzen uns dazu nötigen, den Bereich der inhaltlichen Schlußweisen, die wir für die Überlegungen der Beweistheorie verwenden, weiter zu fassen, als es unserer bisherigen Durchführung des finiten Standpunktes entspricht.

[113] In the context of this discussion, on p. 348/362, a very concise explication of "Begriff der finiten Aussage" is given that reflects faithfully the informal considerations on which I have been reporting.

[114] *Hilbert and Bernays 1939*, p. 357/371. That point is reemphasized after the reductive argument has been completed. The German text there is: Stellen wir uns andererseits auf einen inhaltlichen Standpunkt, von dem aus die formalen Ableitungen in ($\mathbf{Z}$) als Darstellung richtiger inhaltlicher Überlegungen deutbar sind, so ist für diesen auf Grund der festgestellten Beziehung zwischen den Formalismen (Z) und ($\mathbf{Z}$) die Widerspruchsfreiheit des Systems (Z) ersichtlich.

[115] *Hilbert and Bernays 1939*, p. 358/372. The fuller German text is: Jedoch entfernen wir uns mit dem inhaltlichen Folgerungsbegriff total von Hilberts methodischen Gedanken der Beweistheorie, ...

(p. 43) that intuitionism is a proper extension of finitist mathematics (in sharp contrast to the earlier perspective that was discussed in the second half of section 3.3). Bernays expressed that view also in contemporaneous papers and in many later comments, perhaps most dramatically in his article *1967* on Hilbert, where (on p. 502) the above relative consistency proof for Z is seen as the reason for the recognition "that intuitionistic reasoning is not identical with finitist reasoning, contrary to the prevailing views at the time."

The question is raised, whether — in a proof of the consistency of Z — the use of absurdity can be avoided, as well as the appeal to an interpretation of the formalism (viewed in contrast to its direct proof theoretic examination). It is claimed that Gentzen's consistency proof addresses both these issues. After a thorough discussion of the details of the system of ordinal notation (for ordinals less than the first epsilon number) and the justification of the principle of transfinite induction, but only the briefest indication of the structure of Gentzen's proof, the main body of the first edition of the book concludes with some extremely general remarks about the significance of that proof: it provides a perspective for the proof theoretic investigation also of stronger formalisms, when one clearly has to countenance the use of larger and larger ordinals. The volume concludes with the sentence: "If this perspective should prove its value, then Gentzen's consistency proof would open a new phase of proof theory." In this way, it seems, Bernays sees Gentzen's approach as overcoming "the temporary fiasco of proof theory" he discussed in the introduction to volume II and attributed to "... exaggerated methodological demands put on the theory." No explicit final and definitive judgment on the methodologically appropriate character of Gentzen's consistency proof is articulated in the first edition of the book. However, in the introduction to the second edition Bernays states that the transfinite induction principle used in it is "a non-finitist tool."

In the introduction of the first edition and the detailed discussion sketched here, some see an ambiguity in Hilbert and Bernays' view as to whether the extension of the finitist standpoint necessitated by the incompleteness theorems is essentially still the finitist standpoint as articulated in the first two chapters of volume I or whether it is a proper extension compatible with the broader strategic considerations underlying proof theory. I think the ambiguity, if it is there at all, should be resolved in the latter sense; after all, the considerations in Chapter 5.5 come under the heading "Transcending the former methodological standpoint of proof theory. Consistency proofs for the full number theoretic formalism." One just has to distinguish very carefully, as Bernays does, between the two different tasks I described at the very beginning of this section: (i) exploring the extent of finitist mathematics, and (ii) demarcating the appropriate methodological standpoint for proof theory.

However, there is not even a broad demarcation of a new, wider methodological standpoint for proof theory; a reason for this lack is perhaps implicit in the remarks connecting the consistency proof for Z relative to intuitionistic arithmetic with Gentzen's consistency proof (p. 359/372). It is claimed, first of all, that it is

"unsatisfactory from the standpoint of proof theory" to have a consistency proof for Z that "rests mainly on an interpretation of a formalism." It is observed, secondly, that the only method of going beyond the formalism Z has been the formulation of truth definitions: a classical truth definition was given for Z, and the formalization of the consistency proof based on an intuitionistic interpretation would amount to using a truth definition. Thirdly and finally, it is argued that a consistency proof is desirable that rests on "the direct treatment of the formalism itself;" that is seen in analogy to obtaining the consistency of (primitive) recursive arithmetic, where Hilbert and Bernays were not satisfied with the possibility of a finitist interpretation, but rather convinced themselves of the consistency by specific proof theoretic methods. Where in this discussion is even an opening for a broader demarcation?

Bernays, in the "Nachtrag" to his *1930*, reflects on these issues and indicates, in particular, that the epistemological perspective that was underlying proof theory became problematic. Referring to his own essay he writes:

> ... the sharp distinction between what is intuitive and what is not, as it is used in the treatment of the problem of the infinite, apparently cannot be drawn so strictly, and the reflections on the formation of mathematical ideas still need to be worked out in more detail in this respect. Various considerations for this are contained in the following essays.[116]

Some indications of a general direction for philosophical reflections are indeed contained in the essays reprinted in *Bernays 1976*, but also in an essay that is not reprinted there, namely, *Bernays 1954*. There he envisions the appeal to what he calls *sharpened axiomatics* (verschärfte Axiomatik) and, opposing it to existential axiomatics, formulates as a minimal requirement that "the objects [making up the intended model of the theory] are not taken from a domain that is thought as being already given, but are rather constituted by generative processes."[117]

[116] *Bernays 1976*, p. 61. The German text is: ... die scharfe Unterscheidung des Anschaulichen und des Nicht-Anschaulichen, wie sie bei der Behandlung des Problems des Unendlichen angewandt wird, ist anscheined nicht so strikt durchführbar, und die Betrachtung der mathematischen Ideenbildung bedarf wohl in dieser Hinsicht noch der näheren Ausarbeitung. Für eine solche sind in den folgenden Abhandlungen verschiedene Überlegungen enthalten. The "folgenden Abhandlungen" are referring, obviously, to the essays in *Bernays 1976*.

[117] *Bernays 1954*, pp. 11-12. The German text is: Die Mindest-Anforderung an eine verschärfte Axiomatik ist die, dass die Gegenstände nicht einem als vorgängig gedachten Bereich entnommen werden, sondern durch Erzeugungsprozesse konstituiert werden. Bernays continues with a methodologically important remark: Es kann aber dabei die Meinung sein, dass durch diese Erzeugungsprozesse der Umkreis der Gegenstände determiniert ist; bei dieser Auffassung erhält das *tertium non datur* seine Motivierung. In der Tat kann Offenheit eines Bereiches in zweierlei Sinn verstanden werden, einmal nur so, dass die Konstruktionsprozesse über jeden einzelnen Gegenstand hinausführen, und andererseits in dem Sinne, dass der resultierende Bereich überhaupt nicht eine mathematisch bestimmte Mannigfaltigkeit darstellt. Je nachdem die Zahlenreihe in dem erstgenannten oder in dem zweiten Sinne aufgefasst wird, hat man die Anerkennung des *tertium non datur* in bezug auf die Zahlen oder den intuitionistischen Standpunkt. Bei dem finiten Standpunkt kommt noch die Anforderung hinzu, dass die Überlegungen an Hand

There is no indication in that paper or in other writings what kind of generative processes should be considered, and why that particular feature of domains should play a distinctive foundational role. Contemporary proof theoretic investigations give such indications.

6 OUTLOOK BEYOND 1939

What Ackermann formulated in his review of just the first volume of *Grundlagen der Mathematik*, holds even more for the complete two-volume work, namely, that it "is to be viewed in line with the great publications of Frege, Peano, and Russell-Whitehead." In contrast to the other works, these two volumes synthesize contemporaneous research and continue to have a deep effect on research in mathematical logic and proof theory.[118]

Ever since the Gödel-Gentzen reduction of classical arithmetic to its intuitionistic version was obtained in 1932, *foundational reductions* have been pursued. We have been able to establish the consistency of *strong* subsystems of analysis, respectively of *weak* subsystems of set theory. These consistency results have been achieved relative to constructive theories (like theories of constructive number classes and Martin-Löf type theory) or relative to intuitionistic number theory together with transfinite induction for long initial segments of the second number class. Classical analysis as presented in Supplement IV of Hilbert and Bernays can easily be carried out in the theory $(\Pi^1_1 - CA)\restriction$, which is consistent relative to the intuitionistic theory of finite number classes.[119]

While such foundational reductions were being pursued, it was also becoming clear through refined mathematical investigations that most of classical analysis can be carried out in $(\Pi^0_1 - CA)\restriction$ or, what Kreisel called, the theory of arithmetic properties. This theory is like $(\Pi^1_1 - CA)\restriction$, except that the comprehension principle is restricted essentially to arithmetic formulas; it is a conservative extension of (Z) and, thus, actually consistent relative to the system **Z**. The program of Reverse Mathematics, initiated by H. Friedman and S. Simpson, has grown a sort of Kroneckerian offshoot that allows developing substantial parts of analysis and algebra already in conservative extensions of primitive recursive arithmetic and even weaker theories, like F. Ferreiras' feasible analysis. As these conservation results mostly hold also for Π^0_2-sentences one achieves *computational reductions*:

der Betrachtung von endlichen Konfigurationen verlaufen, somit insbesondere Annahmen in der Form allgemeiner Sätze ausgeschlossen werden.

[118] Kreisel's *Survey of proof theory* gives a most interesting, albeit somewhat idiosyncratic perspective. Proof theoretic investigations in Gentzen-style are presented by Schwichtenberg, Feferman in *Buchholz e.a.*, and of course in the textbooks by Schütte, Takeuti, and Pohlers. The constructive understanding of formalisms via (Gödel's Dialectica) interpretation is presented in *Troelstra*, and surveyed in *Avigad and Feferman*. Other constructive approaches were initiated through Martin-Löf's type theory and Myhill's constructive set theory.

[119] $(\Pi^1_1 - CA)\restriction$ is the subsystem of second-order arithmetic with the comprehension principle for formulas of the form $(X)\phi(X)$, where the matrix ϕ is purely arithmetic, and with the induction principle formulated as a second-order axiom, not as a schema for all formulas of the theory.

proofs of such statements yield Skolem functions that lie in particular classes of calculable functions, and proof theoretic tools are used to extract computational information. Kohlenbach has been very successful in obtaining mathematically significant results by this "proof mining."

All of this is in a certain sense internal to proof theory. However, we should consider the stimulus Hilbert's approach and questions provided to contemporaries outside the Hilbert School, for example, to von Neumann, Herbrand, Gödel, Church, and Turing. There is indeed no foundational enterprise with a more profound or more far-reaching effect on the emergence and development of modern mathematical logic. The foundational concerns have led to a dramatic expansion of the reach of mathematical thought. Hilbert would be delighted with the development of mathematical logic in general, and proof theory in particular — in spite of the fact that his Second Problem hasn't been solved yet. He might just note optimistically "Wir müssen wissen, wir werden wissen." and follow this by "Es gibt kein Ignorabimus!" for added emphasis.

Hilbert's specific proposal of mediating between a restricted constructivist and a strong classical position does not work out, when finitism is equated with the methodological stance associated with proof theory. The reductive program that emerged from it provides, however, an important perspective on aspects of mathematical experience and helps us to gain a better understanding of the distinctive character of modern mathematics. It is centrally connected to the starting point of Hilbert's considerations, namely, the foundational opposition of Dedekind and Kronecker. Hilbert's principal goal was to secure mathematics from contradictions — and thus to safeguard creative freedom within mathematics and for contexts of applications. In the Heidelberg talk of 1904 he formulated the creative principle that "in its freest use justifies us in forming ever new notions, with the sole restriction that we avoid a contradiction." As philosophers and mathematicians we should exploit Hilbert's complex insights into the workings of mathematics, instead of keeping him shackled to a narrow foundational position that was taken for exploratory, programmatic reasons.

BIBLIOGRAPHY

The notes of lectures given by Hilbert (and Bernays) I used are available from the Hilbert Nachlass at the Library of the University of Göttingen (SUB mark) or from the Mathematical Institute (marked by MI). Those concerning geometry have been published in volume 1 of *David Hilbert's Lectures on the Foundations of Mathematics and Physics, 1891 to 1933*; volumes 2 and 3 will contain those that are mainly concerned with logic and arithmetic. I refer by *LFMP1* to the first volume of this series. On pp. 607-623 of that volume, one can find a complete listing of Hilbert's lectures. Translations are my own, unless they are explicitly quoted from a published source.

[Ackermann, 1925] W. Ackermann. Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit; Mathematische Annalen 93, 1-36, 1925.

[Ackermann, 1934] W. Ackermann. Review of *Grundlagen der Mathematik*, vol. 1; Fortschritte der Mathematik 60, 17-19, 1924.

[Avigad, 2006] J. Avigad. Methodology and metaphysics in the development of Dedekind's theory of ideals; in: *Ferreiròs and Gray 2006,* 159-186, 2006.

[Avigad and Feferman, 1998] J. Avigad and S. Feferman. Gödel's functional ("Dialectica") interpretation; in: S.R. Buss (ed.), *Handbook of proof theory*; Elsevier, 337-405, 1998.

[Baldus, 1924] R. Baldus. *Formalismus und Intuitionismus in der Mathematik*; G. Braun, Karlsruhe i. B, 1924.

[Behnke, 1976] H. Behnke. Die goldenen Jahre des Mathematischen Seminars der Universität Hamburg; Mitteilungen der Math. Gesellschaft in Hamburg, Bd X, Heft 4, 1976, 225–240, 1976.

[Bernays, 1922] P. Bernays. Über Hilberts Gedanken zur Grundlegung der Mathematik; Jahresberichte DMV 31, 10-19, 1922.

[Bernays, 1927a] P. Bernays. Zusatz zu Hilberts Vortrag über "Die Grundlagen der Mathematik"; Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 6, 89-92, 1927; translated in: *van Heijenoort 1967*, 485-489.

[Bernays, 1927b] P. Bernays. Probleme der theoretischen Logik; in: *Bernays 1976*, 1-16, 1927.

[Bernays, 1930] P. Bernays. Die Philosophie der Mathematik und die Hilbertsche Beweistheorie, 1930; in: *Bernays 1976*, 17-61.

[Bernays, 1935] P. Bernays. Hilberts Untersuchungen über die Grundlagen der Arithmetik. 1935; in: *Hilbert 1935*, 196-216.

[Bernays, 1967] P. Bernays. Hilbert, David; in: *Encyclopedia of Philosophy*, P. Edwards (ed.), Macmillan, New York and London, 496-504, 1967.

[Bernays, 1976] P. Bernays. *Abhandlungen zur Philosophie der Mathematik*; Wissenschaftliche Buchgesellschaft, Darmstadt, 1976.

[Bernstein, 1919] F. Bernstein. Die Mengenlehre Georg Cantors und der Finitismus; Jahresberichte DMV 28, 63-78, 1919.

[Buchholz *et al.*, 1981] W. Buchholz, S. Feferman, W. Pohlers, and W. Sieg. *Iterated inductive definitions and subsystems of analysis: Recent proof-theoretical studies*; Springer Lecture Notes in Mathematics 897, Berlin, Heidelberg, New York, 1981.

[Corry, 1996] L. Corry. *Modern algebra and the rise of mathematical structures*, 1996; Second edition, 2004, Birkhäuser-Verlag, Basel.

[Dedekind, 1872] R. Dedekind. *Stetigkeit und irrationale Zahlen*, 1872; in: *Dedekind 1932*, 315–334 and in: *Dedekind 1969*.

[Dedekind, 1888] R. Dedekind. *Was sind und was sollen die Zahlen*, 1888; in: *Dedekind 1932*, 335-391 and in: *Dedekind 1969*.

[Dedekind, 1889] R. Dedekind. Letter to Keferstein, 1889; in: *van Heijenoort 1967*, 98-103.

[Dedekind, 1932] R. Dedekind. *Gesammelte mathematische Werke*, Dritter Band; R. Fricke, E. Noether, and Ö. Ore (eds.); Vieweg, Braunschweig, 1932.

[Dedekind, 1969] R. Dedekind. *Was sind und was sollen die Zahlen? Stetigkeit und irrationale Zahlen*; Vieweg, 1969.

[Dugac, 1976] P. Dugac. *Richard Dedekind et les fondements des mathématiques*; Vrin, Paris, 1976.

[Ebbinghaus, 2007] H.-D. Ebbinghaus. *Ernst Zermelo: An approach to his life and work*; Springer, 2007.

[Ewald, 1996] W. B. Ewald, ed. *From Kant to Hilbert* - A source book in the foundations of mathematics; two volumes; Oxford University Press, 1996.

[Feferman, 1981] S. Feferman. How we got from there to here; in: *Buchholz e.a. 1981*, 1-15, 1981.

[Feferman, 1988] S. Feferman. Hilbert's Program relativized: proof-theoretical and foundational reductions, Journal of Symbolic Logic, 53 (2), 364-384, 1988.

[Feferman, 1998] S. Feferman. *In the light of logic*; Oxford University Press, 1998.

[Feferman and Sieg, 1981] S. Feferman and W. Sieg. Iterated inductive definitions and subsystems of analysis; in: *Buchholz e.a. 1981*, 16-77, 1981.

[Ferreira, 1994] F. Ferreira. A feasible theory for analysis; Journal of Symbolic Logic, 59, 1001-1011, 1994.

[Ferreira and Ferreira, 2006] F. Ferreira and G. Ferreira. Counting as integration in feasible analysis, Mathematical Logic Quarterly 45, 399-407, 2006.

[Ferreiròs, 1999] J. Ferreiròs. *Labyrinth of thought – A history of set theory and its role in modern mathematics*; Birkhäuser, Basel, 1999; second revised edition, 2007.

[Ferreiròs and Gray, 2006] J. Ferreiròs and J. Gray, eds. *The architecture of modern mathematics – Essays in history and philosophy*; Oxford University Press, 2006.

[Gentzen, 1934] G. Gentzen. Untersuchungen über das logische Schließen, Mathematische Zeitschrift 39, 176-210, 405-431, 1934.
[Gentzen, 1936] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie, Mathematische Annalen 112, 493-565, 1936.
[Gentzen, 1943] G. Gentzen. Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie, Mathematische Annalen 119, 140-161, 1943.
[Gödel, 1931] K. Gödel. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I; Monatshefte für Mathematik und Physik 38, 173-98, 1931.
[Gödel, 1933] K. Gödel. The present situation in the foundations of mathematics, 1933; in: *Gödel 1995*, 36-53.
[Gödel, 1986] K. Gödel. *Collected Works I*; Oxford University Press, 1986.
[Gödel, 1990] K. Gödel. *Collected Works II*; Oxford University Press, 1990.
[Gödel, 1995] K. Gödel. *Collected Works III*; Oxford University Press, 1995.
[Gödel, 2003a] K. Gödel. *Collected Works VI*; Oxford University Press, 2003.
[Gödel, 2003b] K. Gödel. *Collected Works V*; Oxford University Press, 2003.
[Gray, 1992] J. Gray. A 19^{th} century revolution in mathematical ontology; in: D. Gillies (ed.), *Revolutions in mathematics*, Oxford University Press, 226-248, 1992.
[Hallett, 1994] M. Hallett. Hilbert's axiomatic method and the laws of thought; in: A. George (ed.), *Mathematics and Mind*, Oxford University Press, 71-117. 1994.
[Herbrand, 1931] J. Herbrand. Sur la non-contradiction de l'arithmétique; Journal für die reine und angewandte Mathematik 166, 1-8, 1931. Translated in: *van Heijenoort 1967*, 612-628.
[Hertz, 1894] H. Hertz. *Die Prinzipien der Mechanik* (Gesammelte Werke, Band III); Barth, Leipzig, 1894.
[Hilbert, 1889-90*] D. Hilbert. Einführung in das Studium der Mathematik. (SUB 530), 1889–90.
[Hilbert, 1891*] D. Hilbert. Projektive Geometrie. (SUB 535), 1891.
[Hilbert, 1893–94*] D. Hilbert. Einleitung in die Functionentheorie. (SUB 540), 1893–94.
[Hilbert, 1894*] D. Hilbert. Die Grundlagen der Geometrie. (SUB 541), 1894.
[Hilbert, 1894–95*] D. Hilbert. Quadratur des Kreises. (SUB 542), 1894–95.
[Hilbert, 1896–97*] D. Hilbert. Theorie der Functionen einer komplexen Veränderlichen. (MI), 1896–97.
[Hilbert, 1897–98*] D. Hilbert. Zahlbegriff und Quadratur des Kreises. (SUB 549), 1897–98.
[Hilbert, 1898–99*] D. Hilbert. Elemente der Euklidischen Geometrie. (SUB 552), 1898–99.
[Hilbert, 1899] D. Hilbert. *Grundlagen der Geometrie*; Teubner, Leipzig, 1899.
[Hilbert, 1899–1900*] D. Hilbert. Zahlbegriff und Quadratur des Kreises. (SUB 549 extended), 1899–1900.
[Hilbert, 1900] D. Hilbert. Über den Zahlbegriff; Jahresbericht der DMV 8, 180-94, 1900. Reprinted in the third edition of *Hilbert* 1899 , Teubner, 1909, 256-262.
[Hilbert, 1901] D. Hilbert. Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900; Archiv der Mathematik und Physik, 3^{rd} series, 1, 44-63, 213-237; part of the essay is translated in *Ewald 1996*, 1096-1105.
[Hilbert, 1901–02*] D. Hilbert. Zahlbegriff und Quadratur des Kreises. (SUB 549 with "Disposition"), 1901–02.
[Hilbert, 1904*] D. Hilbert. Zahlbegriff und Quadratur des Kreises. (MI), 1904.
[Hilbert, 1905] D. Hilbert. Über die Grundlagen der Logik und der Arithmetik; in: *Verhandlungen des Dritten Internationalen Mathematiker-Kongresses*, pp. 174-85, 1905. Teubner, Leipzig. Reprinted in the third edition of *Hilbert 1899*, Teubner, 1909, pp. 263-279. English translation in: *van Heijenoort 1967*, 129-38.
[Hilbert, 1905a*] D. Hilbert. Logische Prinzipien des mathematischen Denkens. (MI), 1905.
[Hilbert, 1917*] D. Hilbert. Mengenlehre. (MI), 1917.
[Hilbert, 1917–18*] D. Hilbert. Prinzipien der Mathematik. (MI), 1917–18.
[Hilbert, 1918] D. Hilbert. Axiomatisches Denken; Mathematische Annalen 78, 405-415, 1918; reprinted in: *Hilbert 1935*, 146-156. Translated in: *Ewald 1996*, 1105-1115.
[Hilbert, 1920*] D. Hilbert. Probleme der Mathematischen Logik. (MI), 1920.
[Hilbert, 1921*] D. Hilbert. Natur und mathematisches Erkennen; talk given in Copenhagen on March 14, 1921. (SUB 589), 1921.
[Hilbert, 1921–22*] D. Hilbert. Grundlagen der Mathematik. (MI), 1921–22.

[Hilbert, 1922] D. Hilbert. Neubegründung der Mathematik; Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 1, 157-177, 1922; translated in: *Ewald 1996*, 1115-1134.
[Hilbert, 1922–23*] D. Hilbert. Logische Grundlagen der Mathematik. (SUB 567), 1922–23.
[Hilbert, 1923] D. Hilbert. Die logischen Grundlagen der Mathematik; Mathematische Annalen 88, 151-165, 1923; translated in: *Ewald 1996*, 1134-1148.
[Hilbert, 1923–24*] D. Hilbert. Logische Grundlagen der Mathematik. (SUB 567 extended), 1923–24.
[Hilbert, 1926] D. Hilbert. Über das Unendliche; Mathematische Annalen 95, 161-190, 1926; translated in: *van Heijenoort 1967*, 464-479.
[Hilbert, 1930] D. Hilbert. Probleme der Grundlegung der Mathematik; Mathematische Annalen 102, 1-9, 1930.
[Hilbert, 1931] D. Hilbert. Die Grundlegung der elementaren Zahlentheorie; Mathematische Annalen 104, 485-494, 1930; translated in: *Ewald 1996*, 1148-1157.
[Hilbert, 1935] D. Hilbert. *Gesammelte Abhandlungen*, vol. 3, Springer-Verlag, Berlin. 1935.
[Hilbert, 2004] D. Hilbert. *Lectures on the Foundations of Geometry, 1891-1902*, volume 1 of Hilbert's Lectures on the Foundations of Mathematics and Physics, 1891-1933; M. Hallet and U. Majer (eds.), Springer-Verlag, Berlin and Heidelberg, 2004.
[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*; Springer-Verlag, Berlin, 1928.
[Hilbert and Bernays, 1934] D. Hibert and P. Bernays. Grundlagen der Mathematik, Band 1; Springer-Verlag, Berlin, 1934; a second edition was published by Springer in 1968.
[Hilbert and Bernays, 1939] D. Hilbert and P. Bernays. Grundlagen der Mathematik, Band 2, 1939; Springer-Verlag, Berlin; a second edition was published by Springer in 1970.
Almost all the quotations I am using occur in both volumes, and I refer to them by "*Hilbert and Bernays 193m* kkk/lll", where m is either 4 or 9 and kkk is the page number of the first edition, whereas lll is that of the second edition. If a quotation occurs only in the first or only in the second edition the context will make it very clear from which edition I am quoting. And of course, if kkk is the same as lll, I list just one.

[Kneser, 1921–22*] H. Kneser. Private notes of Hilbert 1921–22.
[Kneser, 1922–23*] H. Kneser. Private notes of Hilbert 1922–23.
[Kneser, 1923–24*] H. Kneser. Private notes of Hilbert 1923–24.
[Kohlenbach and Oliva, 2003] U. Kohlenbach and P. Oliva. Proof mining: a systematic way of analysing proofs in mathematics; Proc. Steklov Inst. Math. 242, 136-164, 2003.
[Kreisel, 1968] G. Kreisel. Survey of proof theory; Journal of Symbolic Logic 33, 321-388, 1968.
[Majer, 2006] U. Majer. Hilbert's axiomatic approach to the foundations of science - A failed research program?; in: *Boston Studies in the Philosophy of Science*, vol. 251, 155-184, 2006.
[Mancosu, 1999] P. Mancosu. Between Russell and Hilbert: Behmann on the foundations of mathematics; Bulletin of Symbolic Logic 5, 303-330, 1999.
[Mancosu, 1999a] P. Mancosu. Between Vienna and Berlin: the immediate reception of Gödel's incompleteness theorems; History and Philosophy of Logic 20, 33-45, 1999.
[MacLane, 1935] S. MacLane. Hilbert-Bernays on proof theory; Bulletin of the American Mathematical Society, 162-165, 1935.
[Martin-Löf, 1984] P. Martin-Löf. *Intuitionistic type theory*; Bibliopolis, Naples, 1984.
[Moore, 1997] G. Moore. Hilbert and the emergence of modern mathematical logic; Theoria 12, 65-90, 1997.
[Myhill, 1975] J. Myhill. Constructive set theory; Journal of Symbolic Logic 40, 347-382, 1975.
[Parsons, 2008] C. D. Parsons. *Mathematical thought and its objects*; Cambridge University Press, 2008.
[Peckhaus, 1990] V. Peckhaus. *Hilbertprogramm und kritische Philosophie*; Vandenhoeck & Ruprecht, Göttingen, 1990.
[Peckhaus and Kahle, 2002] V. Peckhaus and R. Kahle. Hilbert's paradox; Historia Mathematica 29, 157-175, 2002.
[Pohlers, 1997] W. Pohlers. *Proof theory – An introduction*; Springer Lecture Notes in Mathematics 1407; Berlin, Heidelberg, New York, 1997.
[Poincaré, 1902] H. Poincaré. Review of *Hilbert 1899*; Bulletin des sciences mathématiques 26, 249-272. 1902.

[Poincaré, 1905] H. Poincaré. Les mathématiques et la logique; Revue de métaphysique et de morale, 13, 815-35, 1905; translated in (Ewald, vol. 2), 1021-38.
[Poincaré, 1906] H. Poincaré. Les mathématiques et la logique; Revue de métaphysique et de morale, 14, 17-34, 1906; translated in (Ewald, vol. 2), 1038-52.
[Poincaré, 1906a] H. Poincaré. Les mathématiques et la logique; Revue de métaphysique et de morale, 14, 294-317, 1906; translated in (Ewald, vol. 2), 1052-71.
[Purkert and Ilgauds, 1987] W. Purkert and H. J. Ilgauds. *Georg Cantor*; Birkhäuser-Verlag, Basel, 1987.
[Ravaglia, 2003] M. Ravaglia. Explicating the finitist standpoint; Ph.D. Thesis, Department of Philosophy, Carnegie Mellon University, 2003.
[Reid, 1970] C. Reid. *Hilbert*; Springer-Verlag, Berlin, Heidelberg, New York, 1970.
[Rowe, 2000] D. Rowe. The calm before the storm: Hilbert's early views on foundations; in: V. Hendricks (ed.), *Proof Theory: History and Philosophical Significance*, Kluwer, 55-94, 2000.
[Schütte, 1977] K. Schütte. *Proof theory*; Springer-Verlag, 1977.
[Schwichtenberg, 1977] H. Schwichtenberg. Proof theory: Some applications of cut-elimination; in: Barwise, J. (ed.), *Handbook of mathematical logic*, North-Holland, 867-895, 1977.
[Sieg, 1988] W. Sieg. Hilbert's program sixty years later; Journal of Symbolic Logic 53, 338-348, 1988.
[Sieg, 1991] W. Sieg. Herbrand analyses; Archive for Mathematical Logic 30, 409-441, 1991.
[Sieg, 1994] W. Sieg. Mechanical procedures and mathematical experience; in: A. George (ed.), *Mathematics and Mind*, Oxford University Press, 71-117, 1994.
[Sieg, 1999] W. Sieg. Hilbert's programs: 1917-1922; Bulletin of Symbolic Logic 5, 1-44, 1999.
[Sieg, 2005] W. Sieg. Only two letters: The correspondence between Herbrand and Gödel; Bulletin of Symbolic Logic 11, 172-184, 2005.
[Sieg and Ravaglia, 2005] W. Sieg and M. Ravaglia. David Hilbert and Paul Bernays, Grundlagen der Mathematik; in: I. Grattan-Guinness (ed.), *Landmark Writings in Western Mathematics, 1640-1940*, Elsevier, 981-999, 2005.
[Sieg and Schlimm, 2005] W. Sieg and D. Schlimm. Dedekind's analysis of number: Systems and axioms; Synthese 147, 121-170, 2005.
[Tait, 2005] W. W. Tait. *The provenance of pure reason – Essays in the philosophy of mathematics and its history*; Oxford University Press, 2005.
[Takeuti, 1987] G. Takeuti. *Proof theory*; second edition; North-Holland, 1987.
[Tapp, 2006] C. Tapp. *An den Grenzen des Endlichen – erkenntnistheoretische, wissenschaftsphilosophische und logikhistorische Perspektiven auf das Hilbertprogramm*; Dissertation, Ludwig-Maximilians-Universität, München, 2006.
[Toepell, 1986] M.-M. Toepell. *Über die Entstehung von David Hilberts "Grundlagen der Geometrie"*; Vandenhoeck & Ruprecht, Göttingen, 1986.
[Torretti, 1984] R. Torretti. *Philosophy of Geometry from Riemann to Poincaré*. D. Reidel Publishing Company, 1984.
[Troelstra, 1973] A. Troelstra, ed. *Metamathematical investigations of intuitionistic arithmetic and analysis*; Springer Lecture Notes in Mathematics 344; Berlin, Heidelberg, New York, 1973.
[van Heijenoort, 1967] J. van Heijenoort, ed. *From Frege to Gödel, a source book in mathematical logic, 1879-1931*; Harvard University Press, Cambridge, 1967.
[von Neumann, 1927] J. von Neumann. Zur Hilbertschen Beweistheorie; Mathematische Zeitschrift 26, 1-46, 1927.
[Weber, 1895] H. Weber. *Lehrbuch der Algebra*; Braunschweig, 1895.
[Weyl, 1918] H. Weyl. *Das Kontinuum*; Verlag von Veit, Leipzig, 1918; English translation by S. Pollard and T. Bole, Dover edition, 1987.
[Weyl, 1927] H. Weyl. Diskussionsbemerkungen zu dem zweiten Hilbertschen Vortrag über die Grundlagen der Mathematik; Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 6, 86-89, 1927; translated in: *van Heijenoort 1967*, 480-484.
[Weyl, 1944] H. Weyl. David Hilbert and his mathematical work; Bulletin of the AMS 50, 612-654, 1944.
[Zach, 1999] R. Zach. Completeness before Post: Bernays, Hilbert, and the development of propositional logic; Bulletin of Symbolic Logic 5, 331-366, 1999.
[Zach, 2003] R. Zach. The practice of finitism: epsilon calculus and consistency proofs in Hilbert's program; Synthese 137, 211-259, 2003.

[Zermelo, 1908] E. Zermelo. Untersuchungen über die Grundlagen der Mengenlehre I; Mathematische Annalen 65, 261-281, 1908; translated in: *van Heijenoort 1967*, 199-215.
[Zermelo, 1930] E. Zermelo. Über Grenzzahlen und Mengenbereiche; Fund. Math. 16, 29-47. 1930; translated in: *Ewald 1996*, 1219-1233.
[Zermelo, 1931] E. Zermelo. Über Stufen der Quantifikation und die Logik des Unendichen; Jahresberichte DMV 41, 85-88, 1931.

HILBERT'S EPSILON CALCULUS AND ITS SUCCESSORS

Barry Hartley Slater

1 INTRODUCTION

In this chapter we shall study the variety of epsilon calculi which have been developed since the days of Hilbert and Bernays. But it is not just the related formal results, and structures which will be of interest. Even in Hilbert and Bernays' *Grundlagen der Mathematik*, epsilon terms originated in the study of certain natural language structures, such as definite descriptions, and they turn out to have a much larger range of linguistic applications, for example in the formalisation of personal pronouns. Up to the second world war, epsilon calculi were studied closely by several mathematicians and set theorists grouped around Hilbert; some developments of this interest have appeared since. But the major concern with these calculi subsequently has been with them either as pure logics to be studied in their own right [Leisenring, 1969], or as applied in the study of certain linguistic structures, and their related philosophical problems [von Heusinger and Egli, 2000]. Thus the validation of syllogisms with forms not considered before, like 'All men are mortal; there is a man; so he is mortal', has become available; and difficulties with paradoxical referring phrases like 'the least number not denotable in less than nineteen syllables' (which contains less than nineteen syllables) have been overcome. Epsilon calculi have therefore emerged, more recently, as 'logic' in the sense this word had before the intrusion of foundational studies in Mathematics, as in Frege's *Grundgesetze*, and Russell and Whitehead's *Principia Mathematica*. In this respect they force a return to the 'logic as language' paradigm, and a move away from the 'logic as calculus' one, as argued for independently by, for instance, Cocchiarella [Cocchiarella, 1988].

Epsilon terms ('epsilon' will be used throughout, although 'sigma', 'tau', 'iota', and 'eta' terms have been used to the same effect, by different writers, as we shall see) were introduced by Hilbert and Bernays to provide a basis on which to define the existential and universal quantifiers of standard predicate logic. There was subsequently the formal development of 'the epsilon substitution method'; see, for example, [Ackermann, 1924; Ackermann, 1940; Kreisel, 1951; Kreisel, 1952; Tait, 1965; Mints, 1989; Mints, 1994; Mints, 1996; Mints, 1997; Zach, 2003]. Equally, Hilbert and Bernays used the epsilon calculus to prove two central metatheorems about the predicate calculus: the First Epsilon Theorem, a relative of Herbrand's Theorem, from which came the development of semantic tableaux,

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

for instance; also the Second Epsilon Theorem, showing that epsilon calculi are conservative extensions of the predicate calculus, which thus provides a kind of completeness theorem for the predicate calculus. Quite short, semantical proofs of the Second Epsilon Theorem are common these days, but Hilbert and Bernays used proof theoretic reasoning — though even that in a more complex way than is now necessary. But while epsilon calculi had these crucial formal functions, and there were other formal developments of related kinds as well, some of which we will trace below, we will not just investigate such aspects, since, in line with more contemporary concerns, we shall also follow the large, philosophical and linguistic discussion of natural language structures upon which epsilon calculi throw a very revealing light.

Appreciation of this larger content and application of epsilon calculi has only developed over a period. Hilbert and Bernays used epsilon terms for the various formal purposes above, but the strict definition of an epsilon calculus, structured and discussed as a full formal logic comparable to the predicate calculus, is only first found in Bourbaki's *Éléments de Mathématique* (though see also [Ackermann, 1937–8]). The classical epsilon calculi, at that time, were also developed without the benefit of any formal semantics, but these came into prominence through the work of Asser in 1957, which was later published in the major text by Leisenring, in 1969. Even in Leisenring, however, readings of epsilon terms in ordinary language were a rarity — a clear mathematical sense in terms of the least number operator was given in some cases, but this, by its nature, could not apply any more generally.

It is part of the history of this period that logic, more widely, turned towards such mathematics, but a study of epsilon calculi forces a re-appraisal of this trend. We shall find that epsilon terms, for instance, generally have a natural language rather than mathematical reading. Even the epsilon terms which refer to numbers have this feature, since 9, for instance, we will see, is identical with *the number of the planets*. But also, even more important, in the move away from mathematics, we shall find there are predicates without a number (c.f. [Bunt, 1985, pp. 262–263]), as was here the case before the planets were formed, and there was just planetary material. For some predications involve mass terms, which denote substances rather than individuals. The reverse point of view, namely that all predicates are count terms, has been implicit in much twentieth century practise, but has not always been made explicit. It comes from the attempt to arithmetise Analysis in the late nineteenth century, and the forces before which led to that, and it resulted at one time in the belief that all predicates determine sets, rather than, for instance, sets or proper classes. Russell's Paradox, of course, disabused most logicians of the former idea.

Within symbolic logic, however, Tarski's semantics has been a major contribution which continues the older, set-theoretic tradition, since Tarski took all predicates to determine sets. And in this tradition we find Hughes and Cresswell, for instance, made one relevant explicit assertion, 'there must be some number which is the number of the planets' [Hughes and Cresswell, 1968, p. 144], whose larger content is only implicit in their account of Tarski's semantics [Hughes and

Cresswell, 1968, pp. 135–136]:

> For predicate variables we proceed as follows: if F is a monadic predicate it is convenient for our purposes to think of the property F as being determined by the set of things that possess it; e.g. if Fx means 'x is human' we can, instead of thinking of F as the abstract property of humanity, think of it in terms of the set of human beings.

We shall find that this is false, so that predicates generally must not be interpreted mathematically, and the related, but much broader question of the relation between formal logic and natural language, as a result, starts to become very pressing. This will happen in several ways. The many attempts to add epsilon terms to Intuitionistic Logic, for example, take us well away from anything which is naturally interpretable, which forces a re-assessment of the viability of this as an alternative formal logic, as we shall see. Simply adding epsilon terms, and their definitional axiom, to Intuitionistic Logic produces Epsilon Intuitionistic Logic, but this is not a conservative extension of Intuitionistic Predicate Logic; and the second epsilon theorem in this context only extends to formulas in prenex normal form in Intuitionistic Predicate Logic [Rasiowa, 1956].

A natural language meaning for epsilon terms was intended even in Hilbert and Bernays' own treatment. For at the end of book 1 of Hilbert and Bernays' *Grundlagen* we find a rival theory of definite descriptions to Russell's, and the later introduction of epsilon terms is directly related to this. In Russell's theory of definite descriptions there are, it will be remembered, three clauses: with 'The King of France is bald' these are 'there is a king of France', 'there is only one king of France' and 'he is bald'. Russell used an iota term to symbolise the definite description, but it is not an individual symbol: it is an 'incomplete' term, as he explained it, since 'The King of France is bald' is taken to have the complex analysis,

$$(\exists x)(Kx.(y)(Ky \supset y = x).Bx),$$

and so it does not have the elementary form 'Bx'. Russell hypothesised that, in addition to the linguistic expressions gaining formalisations by means of his iota terms, there was another, quite distinct class of expressions, which he called 'logically proper names'. Logically proper names would, amongst other things, take the place of the variable in such forms as 'Bx'. Russell suggested that demonstratives might be in this class, but he could give no further formal expression to them.

Hilbert and Bernays themselves introduce a kind of complete symbol, by contrast with Russell, defending what would later be called a 'pre-suppositional theory' of definite descriptions. The first two clauses of Russell's definition,

$$(\exists x)(Kx.(y)(Ky \supset y = x)),$$

are not taken, by pre-suppositionalists, to be part of what is asserted by 'The King of France is bald'; they are, instead, the conditions under which one is allowed to introduce into the language an individual term for 'the King of France', which then satisfies the matrix of the quantificational expression above, and becomes a proper symbol to replace the variable in such expressions as 'Bx'. Hilbert and Bernays still used an iota term for this purpose, although it is quite different from Russell's iota term, since, when it is part of the language, it is equivalent to the related epsilon term [Kneebone, 1963, p. 102]. We shall find that epsilon terms, being complete symbols, are the 'logically proper names' Russell was looking for, and that their natural reading is indeed as forms of demonstratives.

It is at the start of book 2 of the *Grundlagen* that Hilbert and Bernays introduce epsilon terms. They first go on to produce a theory of non-definite descriptions of the same pre-suppositional sort to their theory of definite descriptions. Thus they permit an eta term to be introduced into the language if the first of Russell's conditions is met, '$(\exists x)Kx$'; this term then satisfies the associated matrix, but it is, in general, an individual, pre-suppositional term of the same kind as their iota one. There is a singular difference in certain cases, however, since the pre-supposition of the eta term can be proved conclusively, for certain matrices. Thus we know, for any predicate 'F', that

$$(\exists x)((\exists y)Fy \supset Fx),$$

since this is a theorem of the predicate calculus. The eta term this theorem permits us to introduce is what Hilbert and Bernays call an epsilon term, so we get the epsilon axiom schema

$$(\exists y)Fy \supset F\epsilon xFx.$$

But an epsilon term is very unlike the generality of eta terms, since its introduction is clearly not dependent on any contingent facts about F. It is this which permits completely formal theories using epsilon terms to be developed, because such epsilon terms, unlike Hilbert and Bernays' iota terms, are always defined, and, as the theorem indicates, they refer to exemplars of the property in question. The above predicate calculus theorem, in other words, provides the existence condition for certain objects, which the various epsilon calculi then go on to symbolise reference to, using epsilon terms. Copi has explained the theorem's relation with exemplars very fully [Copi, 1973, p. 110].

Kneebone read epsilon terms as formalising indefinite descriptions [Kneebone, 1963, p. 101], and this idea is commonly also found in the work of his pupil, Priest, although strangely Priest himself has pointed out that reading '$(\exists x)(Gx.Fx)$' as '$G\epsilon xFx$' will not do ([Priest, 1979, p. 6], see also [Slater, 1988b, p. 285]). Hilbert read the epsilon term in the above case 'the first F', which indicates its place in some, otherwise unspecified well-ordering of the F's — for instance, in connection with arithmetical predicates, that generated by the least number operator. The same general idea provided Bernays with a form of the Axiom of Choice, in his

epsilon formulation of Set Theory [Bernays, 1958, Ch VIII]. So 'ϵxFx' is not 'an F'. Moreover, as Copi's discussion makes very clear, it is possible that an epsilon term refers to something which is in fact not F — it does this, of course, if there are no F's at all — and that will lead us to theories of reference which materialised only in the 1960s and later, when reference came to be properly distinguished from attribution. If there are F's then the first F is a chosen one of them; but if there are no F's then 'the first F' must be non-attributive, and so denotes something it cannot connote. It functions like a Millian name, in other words, with no applicable sense. With denotation in this way clearly distinguished from description we can then start to formalise the cross-reference which even Russell needed to link his first two conditions 'There is one and only one king of France' with his further condition 'He is bald'. For, by an extension of the epsilon substitution method, the 'he' in the latter comes to be a pronoun for the same epsilon term as arises in the former — whether or not the former is true. And such anaphoric cross-reference in fact may stretch into and across intensional contexts of the kind Russell was also concerned with, such as

George IV wondered whether the author of *Waverley* was Scott.

For, of course, he was indeed Scott, and we may all now know very well that he was Scott. So we obtain a formalisation for transparency in such locutions.

That puts developed epsilon calculi at variance with Fregean views of intensional contexts — and also the Kripkean semantics which has continued to support Frege in this area. But Fregean intensional logic did not incorporate Millian symbols for individuals, and in particular, as we shall see in detail later, that meant it could not clearly distinguish individuals from their identifying properties. The addition of epsilon terms provides the facility for separating, for instance,

$$s = \epsilon x(y)(Ay \equiv y = x),$$

and

$$(y)(Ay \equiv y = s),$$

and so for isolating the proper object of George IV's thought.

Comparable discriminations are then also available with other types of expression. Thus in higher-order epsilon calculi, epsilon terms provide more appropriate formalisations for propositional descriptions like 'what George IV wondered', and numerical descriptions like 'the number of the planets', which resolves a number of difficulties there have been with other formalisations. But there is no problem about even 'the number of the colours', from an epsilon point of view, even though there is no fixed number of colours: the colours form a continuum, with the result that there is only the amount of colour at a certain place, just like there might be an amount of planetary material. And that bears on several matters to do with the foundations of set theory, not the least being some classic puzzles about infinity.

2 THE EPSILON SUBSTITUTION METHOD

The epsilon substitution method is based on the use of epsilon terms to replace the existential and universal quantifiers [Weyl, 1944; Kreisel, 1964]. Epsilon terms are individual terms of the form 'ϵxFx', where 'ϵ' is a variable binding term operator, and 'x' is a free variable in 'Fx'. There may be other free variables in 'Fx', in which case 'ϵxFx' is simply a functional epsilon term; if there are no other free variables in 'Fx' then 'ϵxFx' is a constant epsilon term. Sometimes only such constant epsilon terms are called 'terms', with the generality of epsilon terms then being 'quasiterms', but we will not keep to that usage here. The main axiom which governs epsilon terms is:

$$Fy \supset F\epsilon xFx,$$

with a free variable 'y', and some appropriate arrangement to avoid clashes of bound variables [Meyer Viol, 1995, p. 8], [Leisenring, 1969, p. 53]. On this basis the existential quantifier can be defined via

$$(\exists x)Fx \equiv F\epsilon xFx,$$

which means that, by suitable introduction of negations,

$$(x)Fx \equiv F\epsilon x\neg Fx.$$

Eliminating quantifiers using these definitions evidently turns all general statements formulable in the language of predicate logic into statements about individuals, which Hilbert and his school saw as replacing objectionable 'infinitistic' elements by tractable 'finitistic' ones.

Hilbert distinguished two kinds of arithmetical statement. 'Real' statements, such as '$2 + 1 = 1 + 2$', are reducible to identities involving computable functions for given numerical arguments. They refer only to finite totalities, and hence are capable of immediate verification, at least in principle, by a mechanical computation in a finite number of steps. 'Ideal' statements, such as '$(x)(x + 1 = 1 + x)$', essentially involve infinite totalities, typically the ranges of unbounded universal and existential quantifiers. These do not obviously allow immediate verification, since a search through an infinite totality is not a mechanical procedure. Hilbert wanted the justification for such general statements to be reduced to proving that their real consequences are true, with the proof of this fact being finitist, i.e. using only computable functions and free-variable properties of their values. The epsilon reduction of the quantifiers enables this to be done, since the proof of an ideal statement, in an epsilon calculus setting, will essentially rest on just a finite set E of 'critical formulas', which are of the form of the main epsilon axiom above. Further epsilon axioms were employed in Hilbert and Bernays' Arithmetic, which we shall look at in closer detail later, but in the present context they can be dealt with quite easily.

Now Hilbert took universal statements to be equivalent to free-variable statements, and so treated them as schemata for their numerical instances. As such they cannot be negated, which meant free-variable statements only had instrumental value, Hilbert believed, and they could always be eliminated in favour of finitistic proofs of their instances. Thus, if the last formula of a proof is a constant combinatorial identity, then replacing all free variables in that proof by arbitrary numerals, and then each epsilon term by an appropriate numerical substitution would immediately yield a variable-free, and so finitist, combinatorial proof of the same identity. So consider, therefore, a substitution S of numerals for the constant epsilon terms in a proof: if all critical formulas in the proof are true under the substitution S then we have a what is called a 'solving solution' for the system E, and so the possibility of a finitist proof of the ideal statement. The same device allows one to extract the numerical content of proofs of existential formulas like $(\exists x)Fx$, for instance, since, with $(\exists x)Fx$ translated as $F\epsilon xFx$, for S to be a solving solution for the proof P of such a formula, then $S(P)$ must just be a proof of Fn where $n = S(\epsilon xFx)$. Similar points arise in other cases, thus, as Skolem showed, for every sentence $(x)(\exists y)Rxy$ provable in Arithmetic there is a function $f(x)$ satisfying $(x)Rxf(x)$ which thus gives the content of the epsilon term $\epsilon yRxy$. The problem invariably reduces to finding a solving solution of the critical formulas in a proof. Hilbert proposed a method of finding such a solution by successive approximation.

We order the critical formulas into a list, and take as the initial substitution one which makes all constant epsilon terms zero. Then, if some substitution is not a solving substitution we modify it by taking the first critical formula in the list which is false under it, and altering the numerical value substituted for the epsilon term in that critical formula. Thus if we have

$$S(Fy) \text{ is true, but } S(F\epsilon xFx) \text{ is false,}$$

then we make S' the same as S except

$$S'(\epsilon xFx) = \text{ the least } n \text{ less than or equal to } S(y) \text{ such that } S(Fn) \text{ is true.}$$

In this way we automatically take care of the further, 'least number' axiom (c.f. [Leisenring, 1969, p. 98]), and generate a sequence of substitutions which successively verify more and more of the critical formulas in the form of the main epsilon axiom. What remains to be shown, therefore, is that this sequence terminates after a finite number of steps in a solving solution. Ackermann in 1940 was able to prove termination for full first order Arithmetic, i.e. pure number theory (see also [Wang, 1962, Ch 14]). A much simpler, though non-effective termination proof for full first order Arithmetic was later outlined by Kreisel [Kreisel, 1951; Kreisel, 1952], see also [Tait, 1965; Mints, 1982].

A proof of the existence of a solving solution is now available for the case of Analysis, i.e. second order Arithmetic, although not in this simple manner. In any such solving solution the substitutions for second-order epsilon terms are, of

course, not numerals; they are taken to be lambda abstracts. Ackermann's early attempt at a proof for this case, in 1924 [Ackermann, 1924], was quickly shown to contain a defect by von Neumann, in 1927 [von Neumann, 1927], although von Neumann was unable to replace it with a proof in the general case. Mints, Tupailo and Buchholz have obtained a complete proof only more recently, in 1996 [Mints *et al.*, 1996; Mints and Tupailo, 1999]. It should be noted that these proofs are not in the natural sense 'finitistic', since they follow Gentzen's use of transfinite induction up to the first epsilon number in his proof of the consistency of Arithmetic. But Hilbert's school came to be satisfied that this still conformed to the basic requirements of their philosophical position (see [Gödel, 1980; Tait, 1981; Zach, 2003], also, for instance, [Tiles, 1991, pp. 120–121]).

Even other central aspects of Hilbert's 'finitism' are questionable in certain ways — his notion of free-variable formula, for instance, and his attitude to Induction [Slater, 1992a]. Detlefsen has also questioned these [Detlefsen, 1986, pp. 46–8] and shown that, in the former case, that faces Hilbert with 'Frege's Problem' [Detlefsen, 1986, pp. 9–11]. Thus Hilbert said [Hilbert, 1967, p. 378]:

> ...we come upon a transfinite proposition when we negate a universal assertion, that is, one that extends to arbitrary numerals. So, for example, the proposition that, if x is a numeral, we must always have
>
> $$x + 1 = 1 + x$$
>
> is from the finitist point of view incapable of being negated. This will become clear for us if we reflect upon the fact that the proposition cannot be interpreted as a combination, formed by means of 'and', of infinitely many numerical equations, but only as a hypothetical judgement that comes to assert something when a numeral is given.

Detlefsen was 'bewildered' by this, finding the reasoning behind it 'perplexing', for it makes out that a generalisation is not a genuine proposition. The question had been raised before by Frege [Frege, 1971, p. 86]:

> From the fact that the pseudo-axioms do not express thoughts it...follows that they cannot be premises of an inference chain. Of course, one really cannot call...groups of audible or visible signs...premises anyway, but only the thoughts expressed by them. Now in the case of the pseudo-axioms, there are no thoughts at all, and consequently no premises. Therefore when it appears that Mr Hilbert nevertheless does use his axioms as premises of inferences and apparently bases proofs on them, these can be inferences and proofs in appearance only.

Indeed, how can one get from a non-proposition to a proposition, when deducing an instance from a generalisation, if generalisations are free-variable formulas, and these are not propositions?

The confusion in Hilbert's mind, I think, is connected with the shift which was in his time under way in the expression of generality: from the simple free-variable, algebraic forms which had been the common currency for several centuries, to the quantified forms, capable of greater complexity, popularised by Frege and the more recent logicians. There is an inconsistency in the double symbolism if it is thought that free-variable expressions express something different, while Fregean predicate language is expressively complete. But, if they are just alternative expressions for generality, then the open formulas which Hilbert thought signified nothing are also finitary — using Hilbert's own elimination strategy. Thus they are not in contrast with clearly finitary expressions like '2 + 1 = 1 + 2'. If the open formula

$$x + 1 = 1 + x,$$

for instance, is equivalent to the quantified generalisation

$$(x)(x + 1 = 1 + x),$$

then it is

$$a + 1 = 1 + a,$$

for $a = \epsilon x \neg(x+1 = 1+x)$. And not only has '$a+1 = 1+a$' straightforwardly all the supposed inferential properties of the free-variable formula, it also can be negated, and can be a proper Fregean premise, or conclusion. Epsilon terms are not like, say, the arbitrary names in some predicate calculus proofs and deductions, which cannot occur outside the calculation; epsilon terms are calculatory instruments and more, since they can also occur in premises and conclusions which may be asserted.

Even in Kreisel's more fine grained, later discussion, the point is not cleared up. Indeed, Kreisel took it that there were three categories of mathematical statement [Kreisel, 1951, pp. 241–242]: there were formulas without variables, as before, and free-variable formulas, but also certain others. The free-variable formulas he took to be still decidable, since, with numerals for the individual variables, and computable functions for the function variables, the result is in the previous category. Hence finitist arguments, he thought, could establish free-variable formulas, and these might then be used as schemata for elementary proofs of their particular cases. Were they then alternative expressions for generality? If they weren't they faced Frege's Problem; but if they were then how were they not just a subclass of Kreisel's third category? For, added to the previous two categories, according to Kreisel, was a third category, namely formulas containing bound variables. These could be put into prenex normal form with matrices which were free-variable formulas, but, unlike with the free-variable formulas themselves, Kreisel did not consider an argument to be finitist if it contained formulas containing bound variables. Nevertheless, it was for the latter that he formulated his 'no-counter-example' interpretation, which was a kind of finitist interpretation.

Kreisel's 'no-counter-example' interpretation was based on the First Epsilon Theorem, which, like Herbrand's Theorem, shows that if a quantified formula is

provable in the predicate calculus then a certain finite disjunction of instances of its matrix is also provable, in a language enriched by certain epsilon terms — i.e. Herbrand functions. Then, in systems of number theory whose formulas are recursively enumerable, there will be a computable function which identifies (through Gödel numberings) the appropriate disjunction for any given formula, and such that if the formula is proved, then from the proof we can discover which disjunct is verifiable, while if the negation of the formula is proved, then for each disjunct we can find a substitution for the variables in it which makes it false. Kreisel gives the example

$$(\exists x)(y)(\exists z)(t)\neg Axyzt$$

[Kreisel, 1951, p. 246] saying that if this is provable then there is a number r, and function f, such that

$$\neg Aryf(y)t,$$

is verifiable (for all y and t), while if the negative of the formula is provable then, for any of the given r and f, there is a y and t which make it false.

Did Kreisel forget that his Herbrand reduction also applies in the simple case above, where

$$(x)(x+1=1+x),$$

is, in fact, entirely equivalent to

$$a+1=1+a?$$

For the specific 'a' in question in this case, namely '$\epsilon x\neg(x+1=1+x)$', is just a Herbrand constant. And not only is this a perfect illustration of the 'no-counter-example' interpretation — because '$\epsilon x\neg(x+1=1+x)$' reads expressly as 'that x for which $x+1=1+x$ does not hold' — the possibility of having a finitist representation of universal statements throws considerable light on how Induction is to be viewed from a finitistic point of view.

Thus it might be thought [Detlefsen, 1986, p. 48] that Hilbert was obliged to see Induction as just a proof-schema showing how finitist proofs of Fn can be obtained, for each particular value of n — each time from $F1$, and $n-1$ other premises of the form '$Fm \supset Fsm$', where sm is the successor of m. For he could hardly have thought that one needed to run through all such proofs before one could establish $(n)Fn$, and thereby the open formula Fn. But there is certainly no obligation on a finitist to see Induction in the other way, whatever the historical specifics about Hilbert's actual position. For, through the epsilon elimination strategy again, the universal statement at the terminus of the Induction proof is equivalent to a certain individual one, namely $F\epsilon n\neg Fn$, and so one of the previous finitist proofs will establish it (c.f. [Tait, 1965, p. 178], [Leisenring, 1969, p. 93]) — even though we do not know in general which one. If $2=\epsilon n\neg Fn$ then the proof that $F2$ is a proof that $F\epsilon n\neg Fn$, and so $(n)Fn$, if $200=\epsilon n\neg Fn$, then the proof that $F200$ is a proof that $(n)Fn$, etc. We shall consider this matter further in Section 5.

Forgetting the finitary equivalents of universal statements is not confined to Hilbert. Dunn and Belnap, for instance, forgot this in their well-known argument about the Omega Rule and the substitutional interpretation of the quantifiers. They said [Dunn and Belnap, 1968, p. 180]:

> For let X be the set consisting of the sentences $F(t'), F(t''), \ldots$, where $t', t'', \ldots$ are all the names in the language, and let A be $(x)Fx$. Clearly then A is a logical consequence of X, but it cannot be deducible from X. For if A were deducible from X it would be deducible from a finite subset of X (by the definition of *deduction*).

But $(x)Fx$ is deducible from $F\epsilon x\neg Fx$, and so from even just one of $F(t')$, $F(t'')$, ..., if epsilon terms are in the language.

Even writers closer to the specific concerns of the epsilon calculus forget this kind of finitary reduction. For it is common to see, for instance, the predicate logic rule of universal quantifier elimination, '$(x)Fx/Fa$', merely reiterated in sequent calculus, and natural deduction treatments of the epsilon calculus (e.g. [Leisenring, 1969, p. 115], [Routley, 1969, p. 131]). But this rule is instead best replaced with two, namely first '$(x)Fx/F\epsilon x\neg Fx$', and then '$F\epsilon x\neg Fx/Fa$', as we shall see. Epsilon terms of the form '$\epsilon x\neg Fx$' Hilbert first called tau terms [Hilbert, 1923; Zach, 2003], and the 'Transfinite Axiom' which Hilbert first employed, namely

$$F\tau xFx \supset Fy,$$

is equivalent, in Classical Logic at least, to the epsilon axiom we have been considering,

$$Fy \supset F\epsilon xFx,$$

(see, for instance, [Bell, 1993a]). The finitary equivalents of existential statements, in connection with the latter, however, are not so commonly forgotten, for the rule '$(\exists x)Fx/F\epsilon xFx$', is obviously reversible, even though realising the consequence, that one can go on to refer anaphorically to the individual thus identified in the existential statement, took quite a while to dawn.

3 SOME EPSILON CALCULI

Bourbaki's epsilon calculus with identity [Bourbaki, 1954, Book 1] is axiomatic, and just adds to an axiomatic base for the propositional calculus the epsilon axiom given above, and two laws of identity. He took *Modus Ponens* as the only primitive rule, but derived several others, such as Conditional Proof, and *Reductio ad Absurdum*. Thus, in effect, we get:

$$(XvX) \supset X,$$
$$X \supset (XvY),$$
$$(XvY) \supset (YvX),$$
$$(X \supset Y) \supset ((ZvX) \supset (ZvY)),$$
$$Fy \supset F\epsilon xFx.$$
$$x = y \supset Fx \equiv Fy,$$
$$(x)(Fx \equiv Gx) \supset \epsilon xFx = \epsilon xGx.$$

Bourbaki, however, worked with axiom schemata, with all the variables syntactic ones, and used tau rather than epsilon to form what we are calling 'epsilon terms'. He also, in places, had a polish prefix notation rather than the normal infix one, and he preferred (theoretically, if not in actual practise) a graphical system of place bonding instead of the usual bound variables, which enabled him to obtain such results as

$$\epsilon xFx = \epsilon yFy.$$

Bourbaki defined the quantifiers in terms of his epsilon symbol as before, and in his system the other usual law of identity, '$x = x$', is derivable.

The immediate application Bourbaki had for his epsilon calculus was in his Set Theory, although he made that, as was customary at the time, into a foundation for all of mathematics. Bourbaki's Set Theory is in the von Neumann-Bernays-Gödel style, although it does not explicitly introduce 'classes' as distinct entities from 'sets'. Like Boolos' plural terms [Boolos, 1975; Boolos, 1984], classes, if they were introduced, would merely provide an alternative way of making statements involving predicates, and so the sole distinction Bourbaki needs is that some but only some predicates determine sets, i.e. are 'collectivisantes'. All the usual axioms of set theory are found in Bourbaki's theory, but he does not have an Axiom of Choice, in any form, since its effects can be generated using the epsilon symbol. A similar point can be made about Bernays' epsilon version of his set and class theory [Bernays, 1958, Ch VIII], in which various versions of the Axiom of Choice are considered.

Some care must be taken over the relation between axioms of choice and the epsilon calculus, however (and see [Moore, 1982, pp. 253–255]). Fraenkel and Bar-Hillel, for instance, realised that the epsilon axiom provides for choice, since, in general, the referent of the epsilon term, for any predicate, is indeterminate. But they thought [Fraenkel and Bar-Hillel, 1958, p. 184] that the epsilon axiom allowed for a single selection only. An epsilon term, however, may contain a free variable, so there is no difficulty about catering for simultaneous choices; what is not thereby established is whether the entities chosen by, say,

$$\epsilon yRxy,$$

form into a set, i.e. whether the associated predicate is collectivisante. In Bourbaki's system this is provable in certain cases because of the form of his axiom schema of replacement [Bourbaki, 1954, Book 1, p66]. But epsilon terms merely

provide functions — Skolem functions, and Herbrand functions — which associate a certain value with given arguments. However, they do so even when there are no entities entering into the appropriate relation, e.g. in the illustrated case, even when, for instance, $\neg(\exists y)Ray$, or $\neg(\exists x)(\exists y)Rxy$ (which is the former case with $a = \epsilon x(\exists y)Rxy$). And they do so whenever the relation is universal — which means that e.g. $\epsilon yRay$ is defined even when $(y)Ray$ — so they are not dependent, in such a case, on there being a universal set. Choice functions are commonly defined just over non-empty sets, but the notion of choice is significantly extended in connection with epsilon terms, because the choice is first defined with respect to predicates, whether or not they are applicable, or applicable universally.

To make the choice extensional Bourbaki's second epsilon axiom is needed:

$$(x)(Fx \equiv Gx) \supset \epsilon xFx = \epsilon xGx,$$

and then, for those predicates which are collectivisantes, the choice for any such predicate is the choice for the associated set. If the first epsilon axiom is a principle of natural reasoning, therefore, any philosophical doubts about the Axiom of Choice, at least within an epsilon calculus setting, must devolve onto this second axiom. And indeed, if the predicates are not collectivisantes then any choice is not a choice from a set.

Furthermore, the second epsilon axiom, it has been claimed, is also crucially involved in separating Intuitionistic Logic from Classical Logic, and so is philosophically interesting on other grounds; indeed we shall inspect it very closely later. But specifically, in the present connection, it can be proved that when both of the epsilon axioms are added to Intuitionistic Propositional Logic then one can obtain the Law of the Excluded Middle in any intuitionistic theory in which there are two things which are provably non-identical. But if the second epsilon axiom is not employed the Law of the Excluded Middle is not obtainable [Goodman and Myhill, 1978; Bell, 1993a].

In fact, adding just the first epsilon axiom to Intuitionistic Predicate Logic enables, for instance, de Morgan's Laws to be obtained [Bell, 1993b; DeVidi, 1995], so even the addition of just the first epsilon axiom is not conservative. Another case is the classical theorem, which we saw before guarantees the introduction of epsilon terms, namely,

$$(\exists x)((\exists y)Fy \supset Fx).$$

This becomes provable when the first epsilon axiom is added, although it is not provable in standard Intuitionistic Predicate Logic. In a natural deduction form, standard Intuitionistic Predicate Logic still has the same predicate rules as Classical Predicate Logic, viz:

$$Aw/(x)Ax, (x)Ax/At,$$
$$At/(\exists x)Ax, ((\exists x)Ax(At/P))/P,$$

where the proper term 'w', in the rule of universal introduction, should not occur free in any undischarged assumption above 'Aw', and the proper term 't',

in the rule of existential elimination, should not occur in 'Ax', in 'P', or in any assumption other than 'At'. But if

$$Fy \supset F\epsilon xFx,$$

then, with *Modus Ponens* we can get

$$Fy/F\epsilon xFx,$$

and so, by existential quantifier elimination

$$(\exists x)Fx/F\epsilon xFx,$$

whence by Conditional Proof

$$(\exists x)Fx \supset F\epsilon xFx,$$

and by existential quantifier introduction

$$(\exists y)((\exists x)Fx \supset Fy).$$

A good deal of technical work has been done, as a result of such proofs, to create epsilon extensions of Intuitionistic Logic which are conservative — see [Smirnov, 1971; Smirnov, 1978; Dragalin, 1974; Mints, 1974; Mints, 1977; Meyer Viol, 1995], for instance. This is presumably in the hope of somehow 'saving' Intuitionistic Logic. But it is doubtful whether these results are of any more than academic interest. Certainly the various authors of these technical works rarely address any of the immediately arising, philosophical questions about the various modifications they have proposed to the first epsilon axiom — centrally, of course, about whether these modifications relate in any way to natural language constructions, and hence their own, and others' reasoning.

Mints has some appropriate criticisms of some other writers [Mints, 1977, p. 323]:

> In a natural deduction system considered by Leivant [Leivant, 1973], formulas cannot contain ϵ-quasiterms. For example, $(x)Px\epsilon yPxy$ will not be a formula. Moreover, rather inconvenient restrictions are imposed on the rule for [Conditional Proof] (similar to the restriction in [Smirnov, 1971]).

But Mints' own, Gentzen-style system itself involves a curious notion, since it requires all the sequents occurring in a proof to be 'meaningful' in a certain formalistic sense. An epsilon term, say '$\epsilon xAxyz$', appearing in an occurrence of some sequent is only 'meaningful', according to Mints [Mints, 1977, p. 318], if the formula '$(z)(y)(\exists x)Axyz$' is either in the antecedent of the sequent, or some sequent below that occurrence. Mints takes his idea from a natural deduction system of Dragalin [Dragalin, 1974], which has, as quantifier rules:

$$Aw/(x)Ax, (x)Ax/At,$$
$$At/(\exists x)Ax, (\exists x)Ax/A\epsilon xAx,$$
$$((x)Bx.(\exists x)Ax)/B\epsilon xAx,$$
$$((\exists x)Ax.B\epsilon xAx)/(\exists x)Bx,$$

where 'w' is a proper term as before, but 't' is now a term other than an epsilon term. So one can see here, in miniature, how the need for '$(\exists x)Ax$' is put in place, and how it makes 'ϵxAx' 'meaningful' — in its presence the epsilon term operates like an ordinary term. But the question which arises is what the meaning of such signs is, in the ordinary sense of 'meaning': i.e. how are they to be read, how are they to be translated into, say, English? Smirnov's quantifier rules are the same as Dragalin's except for the last, the rule of epsilon elimination, which becomes

$$B\epsilon xAx/(\exists x)Ax,$$

and he has a restriction on the use of subsidiary rules: epsilon symbols must not occur in the eliminable assumptions or conclusion of a subsidiary deduction [Smirnov, 1978, p. 45]. But while the result, in this, and every other case above, is undoubtedly a conservative extension of Intuitionistic Predicate Logic, we are faced with a generalisation of Frege's Problem: what do any of the various epsilon terms and rules in these cases correspond to, in valid natural language argument?

Meyer Viol's several constructions of conservative extensions include one resembling Leivant's, which modifies Conditional Proof in complex ways, and leads to awkward complications, like the failure of Contraposition. It also requires epsilon terms to be formally meaningful in a similar way to Mints. Summing up, Meyer Viol says about this construction [Meyer Viol, 1995, p. 105]:

> We cannot move an existential formula to the right of [a deduction sign] if it leaves its proper term on the left unsupported. In fact, unsupported ϵ-terms on the left should not be interpreted as ϵ-terms at all: in order for an ϵ-term to get an ϵ-*interpretation*, its interpretation as a *witness*, it needs a licensing condition, the presence of an introducing formula.

The notion of a witness Meyer Viol draws from Kripke's semantics for Intuitionistic Predicate Logic, and it is through direct modifications of this semantics that he produces other conservative extensions, when approaching the matter non-syntactically. He says [Meyer Viol, 1995, p. 110]:

> Indeed, in semantics as in proof theory, the problem of 'conservative extension' is not unambiguous. More specifically, there are several ways of expanding Kripke models for intuitionistic logic with epsilon terms. A first strategy goes like the classical method...We choose witnessing objects for ϵxFx in every node where an existential formula $(\exists x)Fx$ is forced, while making arbitrary choices for ϵxFx in all other nodes. The striking feature of this semantics is that it is non-persistent. That is,

> formulas involving epsilon terms may change from true to false when going to nodes higher up, where the interpretation of their epsilon terms has shifted. A second strategy maintains persistence, by going partial: epsilon terms are only interpreted at those nodes where their supporting existential formula is forced. This is not enough: we have to stick with the same choice of a witness higher up. Both strategies solve our problem. Both strategies pay a price: either non-persistence, or partiality. Thus multiplicity of solutions, and their logical side effects, seems a general feature of constructive epsilon calculi.

Specifically, one side effect with the first strategy is that even many intuitionistic propositional principles become schematically invalid, although all their epsilon-free instances remain available. And a side effect with the second strategy is that for formulas 'F' which are either atomic or built from atomic ones using just conjunction and the existential quantifier, then

$$F\epsilon xGx,$$

entails

$$(\exists x)(Fx.Gx).$$

Meyer Viol says about the latter entailment [Meyer Viol, 1995, p. 118]: 'Thus intuitionistic base logics can conservatively support stronger epsilon principles than their classical counterparts!' But the question he does not address is whether any of those principles have any *natural language* counterparts.

One point bordering on a natural language justification, however, is to be found in Meyer Viol's proposals for a further system. His point is pertinent to the principles of referential argument in modal contexts, and arises when he discusses how one might avoid the last step in deriving the previous predicate logic existential theorem, i.e. the move from

$$(\exists x)Fx \supset F\epsilon xFx,$$

to

$$(\exists y)((\exists x)Fx \supset Fy).$$

He makes the following remark [Meyer Viol, 1995, pp. 103, 113]:

> ...we can restrict abstraction [existential quantifier introduction] over the ϵ-term...This course is suggested by the analogy with modal predicate logic with non-rigid terms, and, semantically, this analogy is of course a natural one, given Gödel's translation of Intuitionistic Logic into S4. In this case we can restrict [abstraction] to 'non-modal' contexts...

Indeed this is what Fitting actually did when constructing his epsilon calculus system for first-order S4 [Fitting, 1972; Fitting, 1975]. Fitting said about this [Fitting, 1972, p. 103]:

> A natural first attempt at such a formulation would be to add S4 axioms and rules to a classical ϵ-calculus base. This fails, and the reason is not hard to find. If $X(x)$ is a formula with one free variable, x, classically ϵxXx is intended to be the name of a constant such that, if $(\exists x)Xx$ is true, $X\epsilon xXx$ is true. But in a Kripke S4 model, $(\exists x)Xx$ may be true in two possible worlds but yet there may be no single constant, c, such that Xc is true in both worlds. Thus ϵxXx cannot be thought of as the name of a constant in an ϵ-calculus S4. Instead we treat ϵxXx as a function defined on the collection of possible worlds and such that, if $(\exists x)Xx$ is true in some possible world, the value of ϵxXx at that world is a constant, c, such that Xc is true there.

This is in tune not only with Kripke's model theory but also Frege's views about reference in opaque, intensional contexts. We shall find, however, that study of the epsilon calculus substantially changes our view of Frege on this and related matters, because, without epsilon terms, neither he nor Russell had a clear way of separating individuals from their properties, which hindered each of them from getting the logic of referential terms right. As a result, there is, in particular, no reason to restrict abstraction on the above basis, and the further faith which is then restored in the first epsilon axiom shows Intuitionistic Logic is certainly lacking in one basic reasoning principle, even if that does not settle the question over the Law of the Excluded Middle.

Discussions of the philosophical import of Intuitionistic Logic do not commonly investigate the lack of the first epsilon axiom, and have been more concerned with Brouwer's school's historic reasoning, and what can be said either for or against it. But historical studies are not confined to discussing authors just in terms of the concepts they were conscious of, since further researches may well unearth external facts which are still highly relevant. One banal external fact, for instance, is the infrequency of use of double negative forms, even among the originators and students of Intuitionistic Logic who claim double negative forms have a different content from the associated positive forms. 'Double negative form' does not mean, for instance, 'it is not demonstrated that 777 does not occur in the decimal expansion of π', but 'it is not the case that 777 does not occur in the decimal expansion of π', and the former is more commonly used than the latter. If the postulated difference between 'not-not-p' and 'p' is properly a difference between 'not demonstrated that not p' and 'p', then no doubts about the Law of the Excluded Middle have been expressed.

But Gödel's translation of Intuitionistic Logic into S4 [Gödel, 1969] is itself an even more decisive external fact, since it shows that Intuitionistic Logic is properly not an alternative propositional logic, but instead merely, as Meyer Viol realized, a (mis-symbolised) modal logic. As Kneale and Kneale said [Kneale and Kneale, 1962, p. 681]:

> Heyting's calculus...has only been mistaken for an alternative logic because of the intuitionists' unfortunate custom of talking of theorems

> as statements of provability. No doubt there may be theorems about the provability of theorems, but there must be others which are not of this kind although they involve logical notions such as those of conditionality and universality.

Certainly no-one wants to assert '$LpvL\neg p$', or even '$pvL\neg p$', but that does not show in any way that one should not assert '$pv\neg p$'. Even with vague statements, for instance, which independently give rise to questions about the Law of the Excluded Middle, this is so, although to see this one has specifically to realise, as Kneale and Kneale suggest, that somethings are necessarily true without it being demonstrable that they are so. They are then true not for a reason, but just by choice and decision — as formalised in epsilon terms governed just by the first epsilon axiom. For if a man with no hairs is bald, yet a man with 1000 hairs is not bald, then one cannot have $(n)(Bn \supset Bsn)$, where sn is the successor of n. Hence $(\exists n)(Bn.\neg Bsn)$, and the only question is which number '$\epsilon n(Bn.\neg Bsn)$' is to denote. The verbal choice is quite arbitrary, of course, since various precisifications of our usage with 'bald' are possible; but once nominated the number in question is synthetically necessary, to the concept of baldness then defined.

4 THE TWO EPSILON THEOREMS

The epsilon calculus sketched in the *Grundlagen* was not seen as of great practical use as a logic, and the early proofs of the First Epsilon Theorem and Second Epsilon Theorem needed for theoretical, meta-logical purposes were formidably complex. But this was largely because the approach was axiomatic, and with the development of Sequent Calculus, Natural Deduction, and other means of presenting what are essentially the same logics we get progressively more tractable and transparent calculi, and also meta-logical results.

This is maybe shown most pointedly in connection with proofs of the Second Epsilon Theorem, which shows, as before, that the epsilon calculus is a conservative extension of the predicate calculus. For while the original lengthy proof in Hilbert and Bernays is, seemingly, faultless, the somewhat different, but still Hilbert-and-Bernays-style proof in the standard English text by Leisenring [Leisenring, 1969, pp. 69–79] contains an error, which was corrected by Flannagan [Flannagan, 1975], but in whose corrected proof there is a further error, which Ferrari has corrected in two ways [Ferrari, 1987; Ferrari, 1989]. Leisenring also provided a much shorter proof in a Gentzen-style epsilon calculus ([Leisenring, 1969, p. 125], see also [Wessels, 1977, p. 534], which was corrected in [Yashuhara, 1982]), although this depends on the Cut Elimination Theorem for this calculus which is not straightforward. In place of these there is now a five-line semantical proof available ([da Costa, 1980, 137], [Meyer Viol, 1995, p. 33]), and a one-line proof in a Truth Tree setting. Truth Trees also provide a finer insight into other meta-logical matters.

The basic difficulty in proofs of the Second Epsilon Theorem, in the old style, is that, while constant epsilon terms play the same role in the epsilon calculus as

individual symbols play in the predicate calculus, more complex epsilon terms may also be involved in an epsilon calculus proof, since they may be substituted into the axioms. A systematic reduction of this complexity is needed, if it is capable of being eliminated in a proof of a theorem expressible just in predicate calculus language. Hilbert and Bernays [Hilbert and Bernays, 1970, Book 2, p23f] say that one epsilon term ϵxFx is subordinate to another epsilon term ϵyGy if and only if G contains ϵxFx, and a free occurrence of y lies within it. For instance,

$$\epsilon xRxy$$

is subordinate to

$$\epsilon ySy\epsilon xRyx.$$

Hilbert and Bernays then define the rank of an epsilon term to be 1 if there are no epsilon terms subordinate to it, and otherwise to be one greater than the maximal rank of the epsilon terms which are subordinate to it (the ranks of the two formulas above are therefore 1, and 2 respectively). Using the same general definitions Leisenring proves, or, at least, attempts to prove, two theorems [Leisenring, 1969, p. 72f]. These are, first, a rank reduction theorem which shows that epsilon proofs of epsilon-free formulas in which the second epsilon axiom is not used, but in which every term is of rank less than or equal to r, may be replaced by epsilon proofs in which every term is of rank less than or equal to $r - 1$ (this is the theorem where Flannagan found the emendable error); and, second, a theorem proving the eliminability of the second epsilon axiom in proofs of epsilon-free formulas. Together, these two theorems show that if there is an epsilon proof of an epsilon-free formula, then there is such a proof not using the second epsilon axiom, and in which all epsilon terms have rank just 1. Even though such epsilon terms still might contain free variables, replacing those that do with a fixed symbol 'a' (starting with those of maximal length) reduces the proof to one in a form of epsilon calculus Leisenring calls the 'epsilon star' system, in which there are only constant epsilon terms [Leisenring, 1969, p. 66f]. He shows that proofs in this system can be converted to proofs in the predicate calculus, through a replacement of the constant epsilon terms by individual symbols.

Leisenring's sequent epsilon calculus does not admit a full Cut Elimination Theorem, any more than do the sequent epsilon calculi of Maehara [Maehara, 1955], or Curry [Curry, 1963]. Nevertheless Leisenring proves a Cut Elimination Theorem for epsilon-free sequents [Leisenring, 1969, p. 125] which is sufficient to give him a version of the Second Epsilon Theorem as a consequence, since he has shown that to relevant cut-free proofs there correspond proofs in the epsilon star system [Leisenring, 1969, p. 121]. Wessels attempted to construct a sequent epsilon calculus in which a full Cut Elimination Theorem was provable; she tried to do so by simply adding the following epsilon rule to Gentzen's well known system LK [Gentzen, 1934–5]:

$$G, \neg Ba, Da; G, Bt/G, D\epsilon xBx,$$

in which 'a' is a free variable not occurring in the lower sequent, 't' a term (i.e. a free variable or an epsilon term), 'B' a formula, 'G' a sequence, and 'Da' a non-empty sequence in which at least one member formula contains at least one occurrence of 'a' [Wessels, 1978, p. 528]. Yasuhara, correcting her [Yashuhara, 1982, p. 311], replaces this rule with

$$G, \neg B\epsilon xBx, D\epsilon xBx; G, Bt/G, D\epsilon xBx,$$

and then derives the full Cut Elimination Theorem, and the Second Epsilon Theorem model theoretically. Yasuhara uses the semantics of Asser [Asser, 1957], and shows that all provable sequents are valid with respect to this semantics, and that all sequents are either provable without the use of the Cut rule, or are not valid. He then shows that if there are no epsilon terms in a valid final sequent, then it is provable without the use of the epsilon rule, or epsilon terms.

Meyer Viol also uses a somewhat similar, semantical argument to prove the Second Epsilon Theorem, although he does not use Asser's semantics, but instead takes epsilon terms simply to name Skolem functions [Meyer Viol, 1995, p. 24]. He supposes that a formula (more explicitly: a formula without epsilon terms) is not derivable in predicate logic, which means that there is a standard predicate logic countermodel to it. But this countermodel can be expanded with appropriate Skolem functions, to produce a countermodel which does not affect formulas without epsilon terms. But that shows the formula is not derivable in the epsilon calculus. Da Costa [da Costa, 1980] uses a similar proof, but generalises it to a more general set of variable binding term operators besides the epsilon symbol.

But there is a much shorter, one-line proof of the Second Epsilon Theorem, and it is even constructive. We come by that one-line proof of the Second Epsilon Theorem if we modify appropriately the predicate calculus truth trees, as developed, say, by Jeffrey (as in e.g. [Jeffrey, 1967]). Jeffrey uses the standard propositional rules and the rules of quantifier interchange, which do not need to be modified. But he also uses the following rule of existential quantifier elimination, namely

$$(\exists x)Fx/Fa,$$

where 'a' must be new, and a rule of universal quantifier elimination

$$(x)Fx/Fb,$$

where 'b' must be old, unless no other individual terms are available. By reducing closed formulas of the form '$P.\neg C$' to absurdity he then can prove '$P \vdash C$', and validate 'P, so C' in this calculus. Clearly, once we have added epsilon terms to the language, the first of these rules must be changed to

$$(\exists x)Fx/F\epsilon xFx,$$

but also the second rule must be changed to

$$(x)Fx/F\epsilon x\neg Fx,$$
$$F\epsilon x\neg Fx/Fa,$$

(where 'a' is old) to produce an appropriate proof procedure. Steen reads 'ϵxFx' as 'the most F-like thing' [Steen, 1972, p. 162], and the corresponding reading of '$\epsilon x\neg Fx$' then explains why $F\epsilon x\neg Fx$ entails, for every a, Fa, since if the most un-F-like thing is in fact F, then the most plausible counter-example to the generalisation has proven not to be so.

But there is a more important reason why the rule of universal quantifier elimination must be broken up into two parts. For Jeffrey's rules only allow 'limited upward correctness' [Jeffrey, 1967, p. 167], since in order to show that if an initial sentence is false in a valuation so also is one of its conclusions, Jeffrey has to say, with respect to his universal quantifier elimination rule, that the quantification there be limited merely to the universe of discourse of the path. The pair of epsilon calculus rules which replace this rule ensure, by contrast, that there is 'total upwards correctness', since if it is false that everything is F then for sure one of the stated consequences of the universal statement is false, namely the immediate one. The reverse form of improvement is also apparent with the existential quantifier elimination rule. For Jeffrey, with his rule, can only get 'limited downwards correctness' [Jeffrey, 1967, p. 165], since it is not a strict entailment. Thus, in order to show that if an initial sentence is true in a valuation so is one of its conclusions, Jeffrey has to stretch his 'truth' so that it is a matter of being true either in the given valuation, or some nominal variant of it. The corresponding epsilon rule gets round this difficulty by not using names, only descriptions, and by being, therefore, totally downward correct: if there is an F then that F is F, irrespective of what name is used to refer to it. In explication of the predicate calculus rule it might be said: suppose there is an F, well, call that F 'a', then Fa. But the epsilon calculus precedes any such naming ceremony; it gets hold of the more primitive, original way we have of referring to objects, and on which naming so evidently relies: by means of demonstratives, like 'that F'.

It is now clear why the Second Epsilon Theorem immediately follows. This theorem [Leisenring, 1969, p. 79] states that any epsilon calculus deduction between epsilon-free premises and conclusion may be replaced by a predicate calculus deduction between the same formulas. But in axiomatic and other treatments of the epsilon calculus, as we have seen, even if the premises and conclusion are epsilon-free, still epsilon terms, and epsilon terms of any complexity, might still enter the deduction, making the fact that any such deduction can be transformed into, or replaced by a predicate calculus deduction very hard to prove. But the transformation required to get a predicate calculus proof is now quite patent: simply change to new names all epsilon terms introduced in the epsilon calculus quantifier elimination rules. This proof rests on the fact that, given epsilon-free premises and conclusion all epsilon terms entering the epsilon calculus deduction are constant epsilon terms. So we have above a direct proof of the Second Epsilon Theorem by considering epsilon proofs effectively in Leisenring's epsilon star system (see also [Hazen, 1987, pp. 412–413]). Meyer Viol, in his discussion of Arbitrary Object

Theory [Meyer Viol, 1995, p. 49], also demonstrates the computational equivalence between constant epsilon terms and the arbitrary names Fine uses, in his version of predicate logic [Fine, 1985].

Epsilon terms which are non-constant, by contrast, crucially enter the proof of the First Epsilon Theorem, which is a relative of Herbrand's Theorem. The First Epsilon Theorem [Leisenring, 1969, p. 78] states that if C is a provable predicate calculus formula, in prenex normal form, then a finite disjunction of instances of C's matrix is provable in the epsilon calculus. Herbrand's Theorem states that, if C is a provable predicate calculus formula, in prenex normal form, then a certain finite disjunction of instances of C's matrix is provable in propositional logic — given that the language is enriched with Herbrand functions. The epsilon calculus enables us to formulate Herbrand functions, which arise when universal quantifiers are eliminated from formulas. Thus, for instance,

$$(\exists y)(x)\neg Fyx,$$

is equivalent to

$$(\exists y)\neg Fy\epsilon xFyx,$$

where the epsilon term, which contains a variable, is a Herbrand function. Using similar reductions, all universal quantifiers can be eliminated from prenex formulas, and the additional fact that, so to speak, existential quantifiers symbolise disjunctions makes all such formulas equivalent to disjunctions. The exact statement of the additional fact comes through applying König's Lemma, which means that, if there is no open path through a truth tree then there is some finite stage at which there is no open path. Thus we get, in the case above, that if no valuation makes the second formula's negation true, then the tree of its instances must close in a finite length. But the negative is a universal formula, and, under the universal quantifier elimination rule, that means that a finite conjunction of instances of the matrix Fyx reduces to absurdity, by propositional logic reasoning. By de Morgan's Laws, that makes tautologous a finite disjunction of instances of $\neg Fyx$. By generalisation we get the First Epsilon Theorem, and Herbrand's Theorem.

The epsilon calculus, however, can take us beyond Herbrand's Theorem, for one has to beware of the impression Herbrand's theorem may give that existential statements are equivalent to disjunctions. Existential statements, it may seem, are not individual statements, but say instead that one of a group of objects has a certain property, without specifying which. The group of objects is customarily called the 'domain' of the quantification, which has to be detailed, it consequently seems, when setting out the semantics of these matters. But the reductions available within the epsilon calculus require that we do away with such 'domains', and indeed such semantics — except, maybe, for pedagogic, or pragmatic purposes. For the example above is also equivalent to

$$\neg Fa\epsilon zFaz,$$

where

$$a = \epsilon y\neg F\epsilon xFyx,$$

and so why the above disjunction of instances of $\neg Fyx$ is true is because a certain specifiable one of its disjuncts is true. Herbrand's theorem does not show that an existential statement is equivalent to a certain disjunction; it shows that an existential statement is provable if and only if a certain disjunction is provable. But what is ultimately provable is not just a statement about a series of objects, as lack of epsilon terms might lead us to believe, but instead a statement about one object. Indeed the existential statement is provably equivalent to it. This is what allows the epsilon substitution method to proceed; and it is what permits continued anaphoric reference to the same object by reiteration of the epsilon term. The existential quantifier is, as a result, misleadingly thought of if it is taken to relate to a plural universe of individuals: an existentially quantified statement is just another statement about an individual — merely a nameless one.

A comparable point goes for the universal quantifier. For, as we saw when discussing Hilbert's views on Induction, a universal statement is not the conjunction of its instances, even though it implies them. But that does not make it simply a free variable statement, since it is, in fact, equivalent to a certain one of its instances — the one involving the prime putative counterexample to it. Lack of the ability to specify that prime putative counterexample to a universal statement is what leaves predicate logic theorists, like Jeffrey, saying that if such a statement is false then one of its instances is false without a way of ensuring that that instance has been drawn, in the truth tree below it, except by limiting the interpretation of the universal claim just to the universe of discourse of the path. It therefore seems necessary that there be a 'model' for the universal statement which restricts it to a certain 'domain', so that it does not strictly include everything. But in the epsilon calculus the universal quantifier does, invariably, mean everything, and there is no need to specify its range.

Certainly it might be convenient to do so, to aid comprehension in some cases, just as it was convenient to call $\epsilon y \neg F \epsilon x Fyx$, '$a$' immediately above. But the provision of a name, in such circumstances, is only an illustration, and likewise with the nomination of a domain. Indeed, since domains are sets of a kind, there might not be a domain, because the things in question might not be collectivisantes. We shall look at this question more generally now, in connection with the semantics of epsilon terms.

5 THE SEMANTICS OF EPSILON TERMS

Although nowadays the semantics of epsilon terms is commonly provided by means of choice functions defined on sets generally, the first interpretations given to them were purely arithmetical. Hilbert and Bernays developed a formal arithmetic based on the epsilon calculus [Hilbert and Bernays, 1970, Book 2, p85f], c.f. [Leisenring, 1969, p. 92], in which there is another epsilon axiom, as was mentioned before. That axiom can be variously formulated; a convenient form is

$$\epsilon x Ax = st \supset \neg At,$$

where 't' is taken to be a numeral, and 's' a function. This constrains the interpretation of the epsilon symbol so that it can be taken to be a least number operator, although that reading is not strictly forced, even with st taken to be the successor of t, since the axiom then only ensures that no number having the property A immediately precedes $\epsilon x Ax$.

Nevertheless, this axiom is sufficient to establish the principle of Mathematical Induction, in the form

$$(A0.(x)(Ax \supset Asx)) \supset (x)Ax.$$

For assume the reverse, namely

$$A0.(x)(Ax \supset Asx).\neg(x)Ax,$$

and consider what happens when the term '$\epsilon x \neg Ax$' is substituted in

$$t = 0 v t = sn,$$

which is derivable from the other axioms of number theory. If we had

$$\epsilon x \neg Ax = 0,$$

then, since $A0$, we would have $A\epsilon x \neg Ax$. But since,

$$A\epsilon x \neg Ax \equiv (x)Ax,$$

we know that $\neg A\epsilon x \neg Ax$, which means we cannot have $\epsilon x \neg Ax = 0$. Hence

$$\epsilon x \neg Ax = sn,$$

for some n. But from the above axiom

$$\epsilon x \neg Ax = sn \supset An,$$

hence An, although we must also have

$$An \supset Asn,$$

which requires the impossible $A\epsilon x \neg Ax$ again. Hence the further epsilon axiom is sufficient to establish the principle of Induction.

Now, on the least number interpretation, all the axioms of Hilbert and Bernays' system are in fact true in the standard model, but if there was a finitary proof of this we would have a finitary proof of the consistency of the system, against Gödel's results. Ackermann, however, did produce a finitary proof of the consistency of a *restricted arithmetic*, where the least number axiom is weakened so that the variable x does not have a free occurrence in Ax within the scope of an epsilon symbol. That means that in the principle of induction x does not have a free occurrence in Ax within the scope of a quantifier. The consistency proof in this

case proceeds using a form of the epsilon substitution method, replacing each epsilon term in a proof with an appropriate epsilon-free term in such a manner that every formula in the proof converts into a numerically true formula [Leisenring, 1969, pp. 98–99].

But why cannot there be such a proof in the general case? One aspect of Gödel's proof of the impossibility of a finitary proof in general bears closely on the relation between epsilon terms and choice. Put formally, the situation as Gödel described it is as follows. We have, for a certain formula 'Ax' (see, e.g. [Kneebone, 1963, pp. 238–239]):

$$\neg Pr(x, \text{`}(y)Ay\text{'}) \supset (\exists z)Pr(z, \text{`}Ax\text{'}),$$

and

$$Pr(x, \text{`}(y)Ay\text{'}) \supset (\exists z)Pr(z, \text{`}\neg Ax\text{'}),$$

where 'Q' is the Gödel number of formula 'Q', and '$Pr(x, \text{`}Q\text{'})$' says that x is the Gödel number of a proof of that formula. So if

$$Pr(x, \text{`}(y)Ay\text{'}),$$

then

$$(\exists z)Pr(z, \text{`}\neg Ax\text{'}),$$

although also

$$(\exists z)Pr(z, \text{`}Ax\text{'}),$$

since '$(y)Ay$' entails 'Ax'. Hence if the system is consistent we have

$$(x)\neg Pr(x, \text{`}(y)Ay\text{'}),$$

which means that

$$(x)(\exists z)Pr(z, \text{`}Ax\text{'}).$$

But this result is, significantly, about formulas containing Gödel numbers, which are represented by numerals. And while there would also be derivable '$(\exists x)(x = \epsilon y \neg Ay)$', since every epsilon term has a reference, there would not be derivable '$x = \epsilon y \neg Ay$' for any specific numeral 'x', because any such identity could only be based on choice. So we would not have

$$(\exists x)((\exists z)Pr(z, \text{`}Ax\text{'}).(\exists z)Pr(z, \text{`}x = \epsilon y \neg Ay\text{'})),$$

where 'x' ranges over numerals, and while it would be true in the standard model that

$$(\exists x)((\exists z)Pr(z, \text{`}Ax\text{'}).x = \epsilon y \neg Ay),$$

it is precisely the fact that the first part of the latter is about the numeral signs themselves which prevents the transparent substitution of the second part, and

so prevents the universal expression, which states the facts, being derivable. If instead we had

$$(\exists x)\ (\text{there is a proof that } Ax.x = \epsilon y \neg Ay),$$

as we did when considering Hilbert's views on Induction, then we would have a proof in the standard model of the universal conclusion, since transparent substitution would be available. Gödel's proof, in other words, was a piece of metamathematics, and employed as a result a proof predicate — a predicate of (the names of) formulas. But, by contrast, provability in natural language is not a predicate of formulas: in locutions like 'it is provable that p' formulas are not mentioned but used (to describe models), and so provability there is, grammatically, a propositional operator on the formulas, which, as we shall see later, removes any referential opacity.

So now we can amplify the point made in Section 2 about the Omega Rule, and see just what the crucial difference is between humans and formal systems. It lies in our ability to give formal systems interpretations. For Gödel constructed a proposition which was true but unprovable in the sense that while proofs, within the system, of certain formulas 'Ax' were available for every 'x', there was no proof within the system of the formula '$(y)Ay$' — if the system was consistent. Within an epsilon calculus setting, the explanation is that, while there are proofs of 'Ax' for every numeral 'x', there is no proof of '$x = \epsilon y \neg Ay$' for any numeral 'x' — since otherwise there would be a proof of '$(y)Ay$'. Nevertheless, *within the standard model* it is true that $x = \epsilon y \neg Ay$, for some natural number x, and so from the appropriate facts about the standard model, namely that $A0, A1, A2 \ldots$ etc., we can still deduce that $(y)Ay$, since we know that one of the premises is equivalent to it. In short, by choosing the standard model we *identify* $\epsilon y \neg Ay$ with a natural number; that identity is then necessary; but what is necessary is not always provable, i.e. analytically a priori, since sometimes it is just a matter of decision and choice, as we saw in the resolution of vagueness before.

The relation between epsilon terms and choice functions was first set out by Asser, although Asser's semantics for an elementary epsilon calculus with just the first epsilon axiom interprets epsilon terms in this most general case as denoting certain rather complex functions. Meyer Viol, calling this kind of epsilon calculus an 'intensional epsilon calculus', simply takes the epsilon terms to name chosen Skolem functions, as was mentioned briefly before. Skolem functions arise in connection with Skolem's Theorem, which states that, if a formula in prenex normal form is provable in the predicate calculus, then a certain corresponding formula, with the existential quantifiers removed, is provable in a predicate calculus enriched with Skolem function symbols. A much stronger, and non-metalogical version of this theorem is provable in the epsilon calculus from which Skolem's actual theorem follows, since we can get, for instance,

$$(x)(\exists y)Fxy \equiv (x)Fx\epsilon yFxy.$$

Hence if the left hand side is provable in the epsilon calculus the right hand side

is provable there. But the left hand side is provable in the epsilon calculus if it is provable in the predicate calculus, by the Second Epsilon Theorem; and if the right hand side is provable in the epsilon calculus it is provable in a predicate calculus enriched with the appropriate function symbols. So, by generalisation, we get Skolem's original result.

When we add to the above intensional epsilon calculus Bourbaki's second epsilon axiom

$$(x)(Fx \equiv Gx) \supset \epsilon xFx = \epsilon xGx,$$

the interpretation of epsilon terms is commonly said to be extensional, i.e. in terms of sets, although that requires that the predicates be collectivisantes, as when 'Fx' and 'Gx' are explicit set membership statements. The epsilon terms in that case designate choice functions, i.e. functions which select one from a given set [Leisenring, 1969, p. 19], [Meyer Viol, 1995, p. 42]. In the case where there are no F's the selection is taken arbitrarily from the universe at large, although the selection is invariably the same, which means the second axiom then validates Kalish and Montague's rule for improper definite descriptions, in the form

$$\epsilon xFx = \epsilon x\neg(x = x).$$

Kalish and Montague in fact prove a version of the second epsilon axiom in their system [Kalish and Montague, 1964, p. 246]. The same axiom holds in Hermes' system [Hermes, 1965], although there one also finds a third epsilon axiom,

$$\epsilon x\neg(x = x) = \epsilon x(x = x),$$

for which there would seem to be no real justification.

But the second epsilon axiom itself is a rather odd one, as we have noted before. One strange thing about it is that both Leisenring and Meyer Viol forget to state the assumption that the predicates in question must determine sets before their semantics can apply (although see [Leisenring, 1969, p. 103, p108]). Indeed, they merely presume this in their model theory, which is drawn from Asser, since 'ϵxBx' is just modelled by means of a choice from the set of things which in the model are B. The case where there are no things which are B is then catered for in a special clause dealing with the empty set; but there is no allowance for mass terms, for instance, i.e. for the case where there is some B, although no discrete Bs, and so no set of them. With continuous quantities there are no fixed elements to be counted (c.f. [Bunt, 1985, pp. 262–263 in particular]), even though numbers can be associated with them, in terms of multiples of an arbitrary unit.

In fact, even when there is a set of the things in question, the second epsilon axiom, in the form above, does not apply in general, for we must allow for intensional discriminations between properties, as in, for instance, 'There is a red-haired man, and a Caucasian in the room, and they are different'. In such a case, if there were only red-haired Caucasians in the room, then, given the second epsilon axiom, we could not extend the epsilon substitutions to discriminate the two people in question.

But not only that makes the second epsilon axiom debatable, for its addition to the rest of the classical epsilon calculus with identity which we saw in Bourbaki is not conservative (c.f. [Yashuhara, 1982]). In particular we can prove

$$Q\epsilon xPx \equiv Q\epsilon x(PxvPx),$$

by taking 'Fx' as 'Px', and 'Gx' as '$PxvPx$' in the second epsilon axiom, and then using the other identity axiom given before (Leibniz' Law). But this new theorem is even in the language of the intensional epsilon calculus without identity. This may remind us that it is necessary co-extensionality which is the standard criterion for property identity (c.f. [Hughes and Cresswell, 1968, pp. 209–210]), and so it points to the coherence of an intermediate system, which adds just an intensional version of the second epsilon axiom to the above intensional epsilon calculus, viz:

$$L(x)(Fx \equiv Gx) \supset \epsilon xFx = \epsilon xGx.$$

For then, within an appropriate modal logic, namely one which contains the rule of Necessitation, at least all the predicate co-extensionalities which can be proved will engender identities between the associated epsilon terms. So we can get the above new theorem, and all others derivable in the same way.

Moreover the original second epsilon axiom is then provable, when the predicates express explicit set membership. For if necessarily

$$(x)(x \in y \equiv x \in z) \equiv y = z,$$

and

$$y = z \equiv L(y = z),$$

then

$$L(x)(x \in y \equiv x \in z) \equiv (x)(x \in y \equiv x \in z),$$

and so, from the modalised second axiom we get

$$(x)(x \in y \equiv x \in z) \supset \epsilon x(x \in y) = \epsilon x(x \in z).$$

Adding the second epsilon axiom to the original intensional epsilon calculus would therefore seem to be sensible only so long as all the predicates are understood in terms of set membership. This is not uncommon, indeed it is pervasive in the traditional semantics for predicate logic. But, as Russell's Paradox showed, there is a demonstrable 'naivety' about this understanding, unless it is taken just as an illustration, since it presumes all predicates are collectivisantes. If, by contrast, we respect the fact that not all predicates determine sets then we should remain with just the first epsilon axiom, placed in an appropriate modal setting, and with the modalised version of the second epsilon axiom given above. Then we can apply that epsilon calculus to the special case where the predicates express set membership without any further epsilon axiom. The interpretation of epsilon terms is thus

invariably as Skolem functions, though in the special case of the membership of sets, naturally those Skolem functions turn out to be choice functions.

But an epsilon system which is even more revealing of the distinction between predicates and sets, and also the nature of numbers, can be obtained from the arithmetic set out by Bostock in 1974. This deduces Peano's axioms from Frege's definitions of the (exact) numerical quantifiers [Bostock, 1974, p. 21], starting from [Bostock, 1974, p. 10]:

$$(0x)Fx \equiv \neg(\exists x)Fx,$$
$$(1x)Fx \equiv (\exists x)(Fx.(0y)(Fy.\neg(x = y))),$$
$$(2x)Fx \equiv (\exists x)(Fx.(1y)(Fy.\neg(x = y)))\ldots$$

and using quantification over the resulting numeral places. Induction is then deduced from an axiom which fixes the range of the numerical variables, requiring that any positive number bears the ancestral of the successor relation to 0.

One has to be a trifle careful with the definitions of 0 and 1, however, once one remembers the possibility of mass terms. For if there is no gold, then while the number of ingots of gold, or rings of gold, is zero, there is still no number of gold; and even if there is only one bit of gold, that does not mean there is one gold. A number is always a number of Ps, where 'P' is a count term. But we can define a count term in the following way: if 'n' ranges from 2 upwards then a count term F is such that

$$(\exists n)(nx)Fx$$

or

$$((\exists x)(y)(Fy \equiv y = x).M(\exists n)(nx)Fx))$$

or

$$(\neg(\exists x)Fx.M(\exists n)(nx)Fx),$$

and the second and third disjuncts in this definition are what may be abbreviated to '$(1x)Fx$', '$(0x)Fx$'. For a count term is only guaranteed if there can be two or more of the items in question, and so the possibility of this must always be inserted. Also, the restriction in the case of the number 1 is required not just because of the possibility of mass terms, but also because of proper names: 'is Peter', for instance, will hold of just one thing without 'is one Peter' holding.

That means that, allowing 'n' to range from zero upwards, $\epsilon n(ny)Fy$ is then the number of F's, and, by the epsilon definition of the existential quantifier, it is what numbers them (i.e. $([\epsilon n(ny)Fy]x)Fx$), if they have a number (i.e. if $(\exists n)(nx)Fx$). But in the opposite case it is the resultant arbitrariness of the referent of the epsilon term which must be remembered. With the cows in a field, for instance, we can associate a determinate number, but with the beef there we cannot do this without first specifying some unit — pound, kilogram, etc. Thus it is not that there is some *non-finite*, specific number in the opposite case, but merely that there is no determinate number [Tiles, 1989, p. 63]. If $\neg(\exists n)(nx)Fx$, then $\neg((\epsilon n(ny)Fy)x)Fx$, and so, although the number exists, it does not number the F's.

Tiles has pointed out that the distinction between count terms and mass terms is a distinction lacking in Frege's logic [Tiles, 1989, p. 151], c.f. [Wright, 1983, pp. 2–3]:

> ...the basis on which Frege applies the concept of Number to a concept is that of the existence of a one-one correspondence between the extensions of concepts, and it is just not clear whether, within the framework established by Frege, there could be a one-one correspondence between the class of red things and some other class (the class of blue things, for example). This is because the distinction between concepts which do and concepts which do not determine a unit is not reflected in his formal notation; the notation suggests, but does not assert, that the concept of Number will be applicable to concepts without restriction.

But the case where no unit is determined, which defines mass terms, Bostock has separated out, and handles in terms of amounts [Bostock, 1979]. As a result, only in a case where a unit is determined do we have a determinate number of discrete items, and thereby a set of them. Count nouns are the natural characterisation of those predicates which are collectivisantes, and nouns more generally, it might be said, are the natural expression for 'classes'.

But if we are dealing just with predicates which are set membership expressions, then, as before, we can prove a version of the second epsilon axiom in Bostock's language, viz

$$(n)((nx)Fx \equiv (nx)Gx) \supset \epsilon n(nx)Fx = \epsilon n(nx)Gx.$$

So what we give the name '9' to, for instance, might as much be the number of the planets as it is the number of any other, equinumerous set. The natural numbers can thus be defined by means of any series of appropriately sized sets, and are not just identifiable from one such series, as Benacerraf previously showed [Benacerraf, 1965]. And the definition of 9 is possible in terms of what might otherwise seem to be a physical contingency, since the identity

9 is the number of the planets,

i.e.

$$9 = \epsilon n(nx)Px,$$

is necessary, even though the related, predicative form

9 numbers the planets,

i.e.

$$(9x)Px,$$

is contingent. It is contingent that the number of the planets does number the planets, because it is not necessary that $([\epsilon n(ny)Py]x)Px$. Also there might have been a different set of planets, so that, for example, while in fact

$$(x)(x \in p \equiv Px),$$

with p the actual set of planets, and

$$\epsilon n(nx)(x \in p) = \epsilon n(nx)Px = 9,$$

it is still possible that

$$(x)(x \in p' \equiv Px),$$

with p' the set of planets in some other possible world, and

$$\epsilon n(nx)(x \in p') \neq 9.$$

We could not have this further contingency, however, if the unmodalised second epsilon axiom held universally, since that requires that even if one only has contingently

$$(x)(Fx \equiv x \in z),$$

then one can still get

$$\epsilon xFx = \epsilon x(x \in z).$$

Using the modalised second axiom, if one only has contingently

$$(x)(Fx \equiv x \in z),$$

then one cannot get, on that basis,

$$\epsilon xFx = \epsilon x(x \in z),$$

which allows the sets of F's to vary from one possible world to another, while still being sets of F's.

6 DESCRIPTIONS AND IDENTITY

When one begins to investigate the predicative meaning of epsilon terms, it is perhaps significant that Leisenring does not discuss the epsilon substitution method, or the general place of theorems of the epsilon calculus, outside of Hilbert's philosophy of Finitism. He notes the 'formal superiority' of the epsilon calculus, comparing some of its pedagogic features with the comparable ones in the predicate calculus (see also [Routley, 1969; Hazen, 1987]); but he justifies the epsilon calculus as the better logic merely by reference to the Second Epsilon Theorem [Leisenring, 1969, p. 63]. Apparently its main value, in Leisenring's day, was that it could prove all that was provable in the predicate calculus but in a smarter,

and less tedious way. Epsilon terms, for Leisenring, were just clever calculating instruments.

Evidently there is more to the epsilon calculus than this, but until more recent times only the significance and usefulness of the first epsilon axiom itself has been dwelt upon. There are a couple of further theorems within the epsilon calculus, however, which will show its extended range of application: they are about the nature and identity of individuals, as befits a calculus which systematically provides a means of reference to them.

The need to provide logically proper names for individuals only became generally evident some while after Russell's work on the theory of descriptions (see e.g. [Rosser, 1953]). The major difficulty with providing properly referential terms for individuals, in classical predicate logic, is what to do with 'non-denoting' terms, and Quine, following Frege, simply gave them an arbitrary, though specific referent [Marciszewski, 1981, p. 113]. The approach was formalised perhaps most fully by Kalish and Montague [Kalish and Montague, 1964, pp. 242–243], who gave the two rules:

$$(\exists x)(y)(Fy \equiv y = x)/F\iota xFx,$$
$$\neg(\exists x)(y)(Fy \equiv y = x)/\iota xFx = \iota x\neg(x = x),$$

where, in explicitly epsilon terms, we would have

$$\iota xFx = \epsilon x(y)(Fy \equiv y = x).$$

Kalish and Montague were of the opinion, however, that their second rule 'has no intuitive counterpart, simply because ordinary language shuns improper definite descriptions' [Kalish and Montague, 1964, p. 244]. And certainly, in that period, the revelations which Donnellan was to publish [Donnellan, 1966] about non-attributive definite descriptions, were not well known. But ordinary language does not, we now know, avoid non-attributive definite descriptions, although their referents are not as constant as Kalish and Montague's second rule requires. In fact, by being improper their referents are not fixed by semantics at all: like demonstratives the referents of logically proper names are found only in their pragmatic use.

Stalnaker and Thomason were more appropriately liberal with their referential terms [Stalnaker and Thomason, 1968; Thomason and Stalnaker, 1968]. And their referential terms also had to apply, they knew, in every possible world [Thomason and Stalnaker, 1968, p. 363]:

> In contrast with the Russellian analysis, definite descriptions are treated as genuine singular terms; but in general they will not be substance terms [rigid designators]. An expression like ιxPx is assigned a referent which may vary from world to world. If in a given world there is a unique existing individual which has the property corresponding to P, this individual is the referent of ιxPx; otherwise, ιxPx refers to an arbitrarily chosen individual which does not exist in that world.

Stalnaker and Thomason realised that 'A substance term is much like what Russell called a *logically proper name*', but thought that an individual constant might or might not be a substance term, depending on whether it was more like 'Socrates' or 'Miss America' [Thomason and Stalnaker, 1968, p. 362]. It was Stalnaker and Thomason, however, who inspired Fitting's S4 epsilon calculus, mentioned before, in which substance terms were absent (see [Fitting and Mendelsohn, 1998, p. 195]). A fuller coverage of identity and descriptions, in modal and general intensional contexts, is to be found in [Routley *et al.*, 1974], and [Routley, 1977], see also [Hughes and Cresswell, 1968, Ch 11]. But, regretably, the epsilon theory of descriptions presented in the first edition of Hughes and Cresswell's classic text, as an alternative to Russell's theory, does not appear in the second edition [Hughes and Cresswell, 1996]; and there is no treatment of it at all in Fitting and Mendelsohn's recent treatise on first order modal logic [Fitting and Mendelsohn, 1998]. With the Australian thinkers Routley, Meyer and Goddard, however, we find the explicit identification of definite descriptions with epsilon terms, as in

$$\iota x F x = \epsilon x (F x.(y)(F y \supset y = x)),$$

([Goddard and Routley, 1973, p. 558], [Routley, 1980, p. 277], c.f. [Hughes and Cresswell, 1968, p. 203]).

Which theorems in the epsilon calculus are behind these kinds of identification? There is one theorem in particular which demonstrates strikingly the relation between Russell's attributive, and some of Donnellan's non-attributive ideas. For

$$(\exists x)(F x.(y)(F y \supset y = x).G x)$$

is logically equivalent to

$$(\exists x)(F x.(y)(F y \supset y = x)).G a,$$

where $a = \epsilon x(F x.(y)(F y \supset y = x))$. For the latter is equivalent to

$$F a.(y)(F y \supset y = a).G a,$$

which entails the former. But the former is

$$F b.(y)(F y \supset y = b).G b,$$

with $b = \epsilon x(F x.(y)(F y \supset y = x).G x)$, and so entails

$$(\exists x)(F x.(y)(F y \supset y = x)),$$

and

$$F a.(y)(F y \supset y = a).$$

But then, from the uniqueness clause,

$$a = b,$$

and so

$$Ga,$$

making the former entail the latter.

The former expression, as we have seen, encapsulates Russell's Theory of Descriptions, in connection with 'The F is G'; it involves the explicit assertion of the first two clauses, to do with the existence and uniqueness of an F. A presuppositional account like that in Hilbert and Bernays, which was later popularised by Strawson, would not involve the direct assertion of these two clauses: on a pre-suppositional account they form the precondition without which 'the F' cannot be introduced into the language. But both of these accounts forget the use we have for non-attributive definite descriptions. Since Donnellan, we now know that there are no preconditions on the introduction of 'the F'; and 'The F is G', as a result, may always be given a truth value. Hence 'Ga' properly formalises it. If the description is non-attributive, i.e. if the first two clauses of Russell's account are not both true, then the referent of 'The F' is simply up to the speaker to nominate.

But one detail about Donnellan's actual account must be noted at this point. He was originally concerned with definite descriptions which were improper in the sense that they did not uniquely describe what the speaker took to be their referent. And on that understanding the description might still be 'proper' in the above sense — if there still was something to which it uniquely applied. Specifically, Donnellan would originally allow 'the man with martini in his glass' to refer to someone without martini in his glass whether or not there was some unique man with martini in his glass. But someone talking about 'the man with martini in his glass' can be rightly taken to be talking about who this phrase describes, if it does in fact describe someone — Devitt and Bertolet pointed this out in criticism of Donnellan [Devitt, 1974; Bertolet, 1980]. It is this latter part of our linguistic behaviour which the epsilon account of definite descriptions respects, for it permits definite descriptions to be referring terms without being attributive, but only so long as nothing has the description in question. Hence it is not the first quantified statement above, but only, so to speak, the third part of it extracted which makes the remark 'The F is G'.

This becomes plain when we translate the two statements using relative and personal pronouns:

There is one and only one F, which is G,
There is one and only one F; it is G.

For 'it' here is an anaphoric pronoun for 'the (one and only) F', and it still has this reference even if there is no such thing, because that is just a matter of the grammar of the language. Now the uniqueness clause is required for two such statements to be equivalent — without it there would be no equivalence, as we shall see — and that means that the relative pronoun 'which' is not itself equivalent to the personal pronoun 'it'. So it was because Russell's logic could not separate

the (bound) relative pronoun from the (unbound) personal pronoun that it could not formulate the logically proper name for 'it', and instead had to take the whole of the first expression as the meaning of 'The F is G'. Using just the logic derived from Frege, it could not separate out the cross-referential last clause.

But how can something be the one and only F 'if there is no such thing'? That is where a second theorem in the epsilon calculus is required:

$$(Fa.(y)(Fy \supset y = a)) \supset a = \epsilon x(Fx.(y)(Fy \supset y = x)).$$

For the singular thing is that this entailment cannot be reversed, so there is a difference between the left hand side and the right hand side, i.e. between something being alone F, and that thing being the one and only F. We get from the left hand side to the right hand side once we see the left hand side entails

$$(\exists x)(Fx.(y)(Fy \supset y = x)),$$

and so

$$F\epsilon x(Fx.(y)(Fy \supset y = x)).(z)(Fz \supset z = \epsilon x(Fx.(y)(Fy \supset y = x))).$$

By the uniqueness clause we get the right hand side. But if we substitute '$\epsilon x(Fx.(y)(Fy \supset y = x))$' for '$a$' in the whole implication then the right hand side is necessarily true. But the left hand side is then equivalent to

$$(\exists x)(Fx.(y)(Fy \supset y = x)),$$

which is, in general, contingent; hence the implication cannot be logically reversed. As we saw in connection with the number of the planets, having the property of numbering the planets (being alone F) is contingent, but possessing the identity of the number of the planets (being identical with the one and only F) is necessary. The difference is not available in Russell's logic, since only possession of the property can be formalised there. In fact Russell confused the two forms, since possession of an identifying property he formalised using the identity sign

$$a = \iota xFx,$$

making it appear that some, maybe even all identities are contingent. But all proper identities are necessary, and it is merely associated identifying properties which are contingent. Ironically, Frege used a complete term for definite descriptions in his extensional logic, as was mentioned before. But Russell explicitly argued against the arbitrariness of Frege's definition, in the case where there isn't just one F, when setting up his alternative, attributive theory of descriptions, in 'On Denoting'. Had Frege's complete term been more widely used, and, for a start, used in his intensional logic, results like those above might have been more widely known earlier.

Now Hughes and Cresswell appreciated that in addition to 'contingent identities' there were also necessary identities, and differentiated between them as follows [Hughes and Cresswell, 1968, p. 191]:

> Now it is contingent that the man who is in fact the man who lives next door is the man who lives next door, for he might have lived somewhere else; that is *living next door* is a property which belongs contingently, not necessarily, to the man to whom it does belong. And similarly, it is contingent that the man who is in fact the mayor is the mayor; for someone else might have been elected instead. But if we understand [The man who lives next door is the mayor] to mean that the object which (as a matter of contingent fact) possesses the property of being the man who lives next door is identical with the object which (as a matter of contingent fact) possesses the property of being the mayor, then we are understanding it to assert that a certain object (variously described) is identical with itself, and this we need have no qualms about regarding as a necessary truth. This would give us a way of construing identity statements which makes $[(x = y) \supset L(x = y)]$ perfectly acceptable: for whenever $x = y$ is true we can take it as expressing the necessary truth that a certain object is identical with itself.

There is more hangs on this matter, however, than Hughes and Cresswell appreciated. For now that we have the logically proper names, i.e. complete symbols to take the place of the variables in such expressions as '$x = y'$, not only do we see better where the contingency of the properties of such individuals comes from — just the linguistic possibility of improper definite descriptions — we also see, contrariwise, why constant epsilon terms must be rigid — because identities involving such terms are necessary. We shall see later that there is a further reason why constant epsilon terms must be rigid.

Certainly Stalnaker and Thomason did not see the matter in this kind of way. As before, they wanted to say there were two kinds of individual constants: ones like 'Socrates' which can take the place of individual variables, and others like 'Miss America' which cannot. But, as we shall see, in the sense required it is the function 'Miss America in year t' which is in question in the second case, and that is not an individual constant — though it can still take the place of an individual variable. Of course, constant epsilon terms being rigid is not entirely incompatible with Fitting's approach, since Fitting, although he denied there could be rigid terms, at least was clear that he was analysing functional epsilon terms, for which there is still a place. It was Routley, Meyer and Goddard who discussed most fully the possibility of rigid individual terms, and sorted out many of their properties, even while they were not completely convinced about them.

Routley presented several rigid intensional semantics, some objectival and some substitutional [Routley, 1977, pp. 185–186]. One of these semantics, for instance, simply took the first epsilon axiom to hold in any interpretation, and made the value of any epsilon term itself. Using such a rigid semantics, Routley, Meyer and Goddard obtained what has been called 'Routley's Formula', i.e.

$$L(\exists x)Fx \supset (\exists x)LFx,$$

by means of the following transformations. Routley's Formula, on such a semantics, holds for any propositional operator and any predicate, but they illustrated it in the case of necessity, and with 'Fx' as 'x numbers the planets'. Then, with 'ϵxFx' as 'the number of the planets', they said that from

$$L(\exists x)Fx,$$

we can get

$$LF\epsilon xFx,$$

by the epsilon definition of the existential quantifier, and so

$$(\exists x)LFx,$$

by existential generalisation over the term [Routley *et al.*, 1974, p. 308], see also [Hughes and Cresswell, 1968, pp. 197, 204]. But Routley still had doubts about whether a rigid semantics was philosophically unobjectionable [Routley, 1977, p. 186]:

> Rigid semantics tend to clutter up the semantics for enriched systems with ad hoc modelling conditions. More important, rigid semantics, whether substitutional or objectual, are philosophically objectionable. For one thing, they make Vulcan and Hephaestus everywhere indistinguishable though there are intensional claims that hold of one but not of the other. The standard escape from this sort of problem, that of taking proper names like 'Vulcan' as disguised descriptions we have already found wanting...Flexible semantics, which satisfactorily avoid these objections, impose a more objectual interpretation, since, even if [the domain] is construed as the domain of terms, [the value of a term in a world] has to be permitted, in some cases at least, to vary from world to world.

Accordingly, while Routley, Meyer and Goddard defended Routley's Formula, and said, for instance, that there was a number which necessarily numbers the planets, namely *the number of the planets* (np), they thought that this was only *in fact* the same as 9, so that one still could not argue correctly that as L(np numbers the planets), so L(9 numbers the planets). 'For extensional identity does not warrant intersubstitutivity in intensional frames' [Routley *et al.*, 1974, p. 309].

This means that Routley, Meyer and Goddard denied the necessity of identity, but if, instead, one wants to use '=' to embrace it, one merely again distinguishes '$9 = \epsilon n(nx)Px$', from '$(9x)Px$', and from '$9 = \epsilon n(nx)(x \in p)$', as we shall see in even more detail shortly. Vulcan then is without doubt Hephaestos, even though, for instance, stories about different numbers of gods might be told in Greek mythology as opposed to Roman mythology, putting the craftsman god in different contexts.

But to have the rigidity fully clear, the formal objections Routley had to his rigid semantics above must also be met. Routley viewed the imposition of the first

epsilon axiom in any model as 'ad hoc'. He therefore did not see it as logically true, i.e. as a law applicable universally, determining what is a possible model. Instead he thought of it as being something which itself is applicable in certain models rather than others. Certainly *the formula which expresses the first epsilon axiom* might be given a variety of interpretations in different worlds. But, as we shall shortly see, rigidity requires that we move away from that meta-linguistic way of thinking. Mention of the formula has to be eliminated, to be replaced by its use. It is the first epsilon axiom formula *interpreted a certain, fixed way* which necessarily is true in every model.

7 RIGID EPSILON TERMS

I will now discuss some further work, which is a part of this history, but which in particular, I believe, has helped us to see just why reference in modal and general intensional constructions must be rigid. I will explain how it relates to the other formal, epsilon calculus work in this area, mentioned before, by Routley, Meyer and Goddard, and how this intersects with the views of Fitting, also Stalnaker and Thomason. It involves a re-ordering of our views on semantics, as has been evident in a few places already, and it begins outside the area of predicate logic altogether, in a review of the standard semantics of propositional logic.

When one thinks of 'semantics' one thinks of the valuation of formulas, but it is only since the 1920s that this kind of study has supplemented the traditional kind of proof theory, which before that had dominated Logic. In the truth tables that Post and Wittgenstein introduced, for instance, it may indeed seem we get a quite different thing from any 'proof theory'. Proof Theory we might, at one time, have associated with axiomatic procedures, for one thing, but certainly, in all its forms, nowadays, with 'the object language'. The theory of truth which Tarski developed depends centrally upon the separation of object and meta-languages — and in fact we may seem obliged to make such a discrimination because of the threat of contradictions from semantic paradoxes, if there was, by contrast, 'semantic closure' (c.f. [Priest, 1984]). Truth may thus appear to be inescapably a meta-concept, and certainly to be a predicate of sentences. On the other hand, because of Tarski's Theorem, we know that what is true, in Tarski's metalinguistic sense, is not representable.

There is a quite different way of looking at these things, however, which is deliberately non-Tarskian, and which others have followed (see [Prior, 1971, Ch 7], and, for instance, [Sayward, 1987]). This involves, at the propositional level, taking 'it is true that' as the basic truth locution. This is not a meta-predicate of object level sentences, but an object level propositional operator, and its manipulations in truth tables are then just another form of proof procedure — any completeness and soundness results in connection with it merely show that there are two ways of arriving at exactly the same theorems. Use of truth tables proves things, as does use of truth trees; the proofs are not axiomatic ones, but that is another issue. In fact, if there are proofs using the propositional operator notion of truth then this

notion is eliminable, producing immediately a non-'semantical' proof procedure. For operator truth is definable, but eliminable, because of the necessity of

$$p \equiv \text{ it is true that } p,$$

since 'it is true that' is the null, identity modality.

The Liar Paradox is then avoided because self-reference is not obtainable with propositional operators [Slater, 1992c]. It might be true or false of a certain subject s that it is P, and so, in particular, that it is false might be true or false of the thought that *p*. But for that it is false to be true or false *of itself* would be for it to be true or false of the thought that it is false, i.e. something whose expression still contains a waiting pronoun. Can it be said of something without a waiting pronoun, such as the sentence 'The sentence at the top of the page is false', that it is true or false of anything — in particular that it is true or false of itself, if it is the sentence at the top of the page? No: for there is no place waiting in that sentence for the name of any thing it might be true or false of to be inserted. *That it is false* is the sort of thing which might be true or false of something — the sentence at the top of the page, and even that sentence when it is the sentence previously given — but that is a possibility only because 'that it is false' has a space waiting to be filled in the appropriate way. And the presence of that space is just what shows appropriate self-reference is not obtainable.

In fact, that it is false *is* false of the illustrated sentence (no matter where it might be), because sentences are neither true nor false. What might be true or false is that *s* is P, but 'that s is *P*' is not a sentence, it is a noun phrase — a demonstrative referring phrase to a thought. The use of such referring phrases most forcefully appears if the sentence at the top of the page is a Strengthened Liar, like 'The sentence at the top of the page is false, or neither true nor false'. Then — since no sentence at the top of the page is either true or false, and so any there is either false, or neither true nor false — wouldn't we, contradictorily, have to say that the sentence actually there was true, since it states the facts? No: because what is true is simply that the sentence at the top of the page is false or neither true nor false, not 'the sentence at the top of the page is false or neither true nor false'. Hence there is no problem with Liars, or Strengthened Liars; and we shall see this point again, in even more detail, later.

Propositional operators include 'it is provable that', which is distinct from Gödel's provability predicate, as Gödel himself pointed out [Gödel, 1980]. They also include 'it is necessary that' and 'it is believed that', for instance, and the attempt to see these, along with 'it is true that', as metalinguistic predicates was a general feature of the philosophy of logic in the early decades of the last century, for instance in Quine's treatment of modality and intensionality. On that approach, someone would be believing that the Morning Star is in the sky, but the Evening Star is not, if, respectively, they did, and did not assent to the two associated sentences. After Montague's and Thomason's work in the 1960's and 1970's [Montague, 1963; Thomason, 1977; Thomason, 1980] we are now more convinced that 'indirect discourse is not quotational', as Thomason put it, but the

earlier generation considered only what words the agent would use.

The change requires seeing the subordinate sentence in indirect speech as not in the words of the agent, but instead in the words of the reporter. That is indeed the core of what reported speech is about — putting something into one's own words rather than quoting them from another source. Thus someone may say

Celia believed a woman was in the room, but it was the man there.

contains a transparent intensional locution, with the same object 'inside' the belief as 'outside' in the room. Here is where the rigid constant epsilon terms come in, in symbolising the cross-reference:

$$Bc(\exists x)(Wx.Rx).(y)((My.Ry) \equiv y = \epsilon x Bc(Wx.Rx)).$$

But to get a full grasp of the matter, we must parllel in the general case the shift from meta-predicate to object language propositional operator which we saw above with truth.

But to get a full grasp of the matter, we must parallel in the general case the shift from meta-predicate to object language propositional operator which we saw above with truth.

Routley, Meyer and Goddard, it will be remembered, saw that a rigid semantics required treating the 'x' in such expressions as '$BcWx$' as open for quantification. They derived, as before, 'Routley's Formula', which in the case of necessity is

$$L(\exists x)Fx \supset (\exists x)LFx.$$

But now we can see, in a better way than Routley, Meyer and Goddard, how the number of the planets provides a clear illustration of the validity of this formula. The fact that some concepts may not have a number is crucially what must be appreciated, but also it must be remembered how the referential, and attributive senses of individual terms are distinguished. For, with 'Fx' as '$x = \epsilon n(ny)Py$' the entailment holds because identities are necessary, making both sides of the formula true. But if 'Fx' is '$(xy)Py$' this is contingent, since, as before, it is contingent that there is a number of planets, instead of just planetary material; and that makes both sides of the formula false.

Hughes and Cresswell argue against Routley's Formula, saying [Hughes and Cresswell, 1968, p. 144]:

> ...let $[Fx]$ be 'x is the number of the planets'. Then the antecedent is true, for there must be some number which is the number of the planets (even if there were no planets at all there would still be such a number, viz. 0): but the consequent is false, for since it is a contingent matter how many planets there are, there is no number which *must* be the number of the planets.

But this overlooks mass terms, which do not involve a unit without some arbitrary nomination, with the consequence, as we have seen, that any number associated with an amount does not *number the amount*, being related instead to the chosen unit. The attempt to see continua in terms of a fixed number has been central to the development of Analysis, and Set Theory, and it has also influenced advanced logic as well, supporting the set-theoretic semantics of predicate logic. Certainly if the term involved is a count noun then there will be a number (maybe 0) of the F's then determined. But the number associated with some F does not number any Fs, with the result that the antecedent of Routley's Formula is not necessarily true.

Quine, of course, considered the example of the number of the planets in his famous argument against the possibility of quantification into modal contexts. He said [Quine, 1960, pp. 195–197]:

> If for the sake of argument we accept the term 'analytic' as predicable of sentences (hence as attachable predicatively to quotations or other singular terms designating sentences), then 'necessarily' amounts to 'is analytic' plus an antecedent pair of quotation marks. For example, the sentence:
>
> (1) Necessarily 9 is greater than 4
>
> is explained thus:
>
> (2) '9 is greater than 4' is analytic...
>
> So suppose (1) explained as in (2). Why, one may ask, should we preserve the operatorial form as of (1), and therewith modal logic, instead of just leaving matters as in (2)? An apparent advantage is the possibility of quantifying into modal positions; for we know we cannot quantify into quotation, and (2) uses quotation...
>
> But is it more legitimate to quantify into modal positions than into quotation? For consider (1) even without regard to (2); surely, on any plausible interpretation, (1) is true and this is false:
>
> (3) Necessarily the number of major planets is greater than 4.
>
> Since 9 = the number of major planets, we can conclude that the position of '9' in (1) is not purely referential and hence that the necessity operator is opaque.

But Quine fails to distinguish the referential, and necessary 'the number of the major planets is greater than 4' from the predicative, and contingent 'There are more than 4 major planets'. If

$$9 = \epsilon n(ny)Py,$$

then

$$\epsilon n(ny)Py \text{ is greater than } 4,$$

but it does not follow that

$$(\exists n)((ny)Py.n \text{ is greater than } 4).$$

Substitution of identicals in (1), therefore, does yield (3), in its correct interpretation, even though it is not necessary that there are more than 4 major planets. Comparable issues arise with respect to other putative counterexamples to substitutivity, and Routley's Formula, see [Slater, 1990].

But how does the 'x' in such a form as 'LFx' come to be open for quantification? Fitting and Mendelsohn only briefly discuss rigid designators in their recent text book [Fitting and Mendelsohn, 1998, pp. 217–219], and their work stays well within the meta-linguistic paradigm, with no mention of propositional operators at all. In particular that leaves them not only with Frege's puzzle about Hesperus being Phosphorus, but also a worry about whether everyone can be said to know about this. Certainly not everybody can be expected to know that the sentence 'Hesperus is Phosphorus' expresses a necessary truth; but, as two-dimensional modal logic has made clear, that is quite different from knowing that Hesperus is Phosphorus. The propositional operator expression involves the use, not the mention of the given sentence.

In traditional, meta-linguistic modal semantics (see [Hughes and Cresswell, 1968, p. assim]) one finds formulas like

$$V(Fx, i) = 1$$

which are supposed to say that the valuation put on the formula 'Fx' is 1, in world i. But there should be quotation marks around the 'Fx' in this formula, if that is the intention. By a common convention, they are omitted. To move over to the alternative point of view, one simply reads '$V(Fx, i) = 1$' as it stands, in which case it becomes the propositional operator form 'it would be true in world i that Fx'. The 'Fx' is then in reported speech rather than quoted speech, and so the 'x' is in the words of the speaker, and the meta/object distinction has been discarded. Any variable inside the subordinate sentence can then be quantified over, just like any variable outside it, so there is unproblematic 'quantifying in', and all standard predicate calculus, and identity rules apply. The epsilon term 'ϵxFx' is then rigid because, quite generally,

$$p \equiv [V(p, a) = 1],$$

where a is the actual world, and therefore, necessarily

$$(x)(Fx \equiv [V(Fx, a) = 1]),$$

which gives

$$\epsilon xFx = \epsilon x[V(Fx, a) = 1],$$

by the second epsilon axiom. More specifically, remembering the quotation from Fitting before, suppose that in every world there are F's, i.e.

$$A : (i)[V((\exists x)Fx, i) = 1],$$

then we know, by the necessary equivalence produced by the epsilon definition of the existential quantifier, that

$$B : (i)[V(F\epsilon xFx, i) = 1],$$

and that means that

$$C : (i)(\exists x)[V(Fx, i) = 1],$$

i.e. that in every world there is an object such that it is F. But that means, using the definition of the quantifier again, that

$$D : (i)[V(F\epsilon x[V(Fx, i) = 1], i) = 1].$$

Now from D we cannot obtain

$$E : (\exists y)(i)[V(Fy, i) = 1],$$

since '$\epsilon x[V(Fx, i) = 1]$' is not free for 'y' in '$(i)[V(Fy, i) = 1]$'. But we can get E from B, since 'ϵxFx' is a constant, and hence is obviously free for 'y' in '$(i)[V(Fy, i) = 1]$'. Nevertheless, if F is individuating, i.e. contains a uniqueness clause, then from B and D we can get

$$(i)[V(\epsilon xFx = \epsilon x[V(Fx, i) = 1], i) = 1],$$

and so prove that, in all worlds, the F in fact is identical with the F in that world. The respective objects are not thereby identical *tout court*, but their passing identity can make it seem that what goes for one can hardly be separated from what must go for the other. Yet if we think in terms of the functional epsilon term there will be opacity, while with the now properly identified constant epsilon term there clearly is transparency. And that holds even though, respectively, in each world, they refer to the same thing.

A simple example illustrating this involves the winner of a race, and the runners which might have won in other circumstances [Slater, 1988c, pp. 103–104], see also [Slater, 1986a; Slater, 1988a; Slater, 1989; Slater, 1992b; Slater, 1994a; Slater, 1994b]. If the actual winner is Tom — so he is ϵxFx — but it is supposed the winner is Dick — so he is $\epsilon x[V(Fx, i) = 1]$, for the specific 'i' — then what would have to be true for those circumstances to obtain would be for Tom to be Dick, i.e.

$$\epsilon xFx = \epsilon x[V(Fx, i) = 1].$$

Tom is not in fact Dick, but that does not mean they cannot be identical in other worlds (c.f. [Hughes and Cresswell, 1968, p. 190]). Certainly if there were several

people variously thought to have been the winner, those other winners in the various other supposed circumstances would not provide a constant c such that Fc is true in all worlds. But that is because those other winners are functions of the other possible worlds — they are $\epsilon x[V(Fx, i) = 1]$ for other 'i's. It still remains that there is a constant, c, such that Fc is true in all worlds — ϵxFx. Moreover, that c is not an 'intensional object': Tom is just a straightforward extensional object, the actual winner.

Routley, Meyer and Goddard did not appreciate this last point, wanting, in their rigid semantics, the reference of individual terms to be 'intensional objects' ([Goddard and Routley, 1973, p. 561], [Routley *et al.*, 1974, p. 309], see also [Hughes and Cresswell, 1968, p. 197]). But see [Slater, 1988c], again. And what also might have misled Stalnaker and Thomason is that the temporal functionality of 'Miss America', even if it is recognised, is of another sort. For if this year's Miss America is last year's Miss America, that is only one thing being identical with itself, unlike with Tom and Dick. Also there is nothing which can make this year's Miss America to be identical with last year's different Miss America, in the way that the counterfactuality of the situation with the winners forces two actually non-identical things to be the same. Other worlds are categorically different from other times, and other spaces, i.e., in general, other possible worlds must not been seen 'Realistically'.

Hughes and Cresswell describe the model which is misleading all parties [Hughes and Cresswell, 1968, p. 197], and see also [Hughes and Cresswell, 1996, p. 332]:

> Consider, for example, the expression 'the top card in the pack' as it occurs in the rules of a card game. The rules may, without ambiguity, specify that at a certain point in the play the top card is to be dealt to a certain player; yet on one occasion the top card may be the Ace of Spades and on another it may be the Queen of Hearts. Thus 'the top card in the pack' does not designate any particular card (individual piece of pasteboard), except in the context of a particular state of the pack; yet we can in one sense think of it as standing for a single object, contrasted with *the bottom card in the pack*, and so forth...Such 'objects' are often called 'intensional objects', and...for a logic in which...the individual variables range over intensional objects... $L(\exists x)Fx \supset (\exists x)LFx$ would be valid...

But it is *the top card in the pack at one point in the play* which is the Ace of Spades, and *the top card in the pack at another point in the play* which is the Queen of Hearts. So the logic of such expressions requires a formalisation using another variable: say 'Fxt' means 'x is the top card in the pack at time t'. But 'at every time there is some top card' then gets symbolised,

$$(t)(\exists x)Fxt,$$

and while this is

$$(t)F(\epsilon xFxt)t,$$

we cannot use existential generalisation to get

$$(\exists x)(t)Fxt.$$

since

$$\epsilon xFxt$$

is not free for 'y' in

$$(t)Fyt.$$

Hence there are no 'intensional objects' as Hughes and Cresswell, and the others, conceived them.

8 THE EPSILON CALCULUS' PROBLEMATIC

Realism has had a pernicious influence, and delayed a proper construal of many of these things. For, if you look at remarks like "the same object is 'inside' the belief as 'outside' in the room" ontologically, then Realistic views about the mind — as another space or time, where counterfactuality can get no proper purchase — can start to interfere, and make it seem that only some counterpart of the outside object can be inside the belief at the same time. Edelberg is one prominent theorist who is misled in this way (as, e.g. in [Edelberg, 1986; Edelberg, 1995]); for, lacking a proper account of identity, he can only secure 'counterparts' of an object in another place, since the 'object' is identified with its contingent properties in this place or that, and so must be different if those properties differ. He says [Edelberg, 1995, p. 316]:

> The semantical theory developed here makes truth a relation between sentences and theories, or belief systems...In the framework proposed here, the theories or belief systems occupying one argument of the truth relation will have three crucially important features. First, each such theory will have an ontology, a set of *objects* that existaccording to the theory. Second, although no object will ever grace the ontologies of two different theories, objects from distinct theories can be *Counterparts*. The Venus of your theory can be a counterpart of the Venus of mine, even if we disagree about its mass. Third, objects will have an unconventional *mereological structure*: objects playing distinct theoretical roles in one theory will sometimes combine in another to form a third, composite object playing all the theoretical roles of the former objects, and containing them — or rather their counterparts — as subobjects. For instance, the Venus of modern astronomical theory will be composite, containing as subobjects counterparts of the Hesperus and Phosphorus of ancient astronomy.

But if what Edelberg is talking about is that $BeMv$ but $ByM'v$, i.e. that him and another might think Venus has different masses, then there are no different

objects of thought. On the other hand there is no difficulty, instead, in our talking about, say, *what Edelberg or some other* believes rises in the morning, or sets in the evening ($\epsilon xBeRMx, \epsilon xBySEx$), just as we talked about $\epsilon x\{V(Fx, i) = 1\}$ before. And maybe, indeed, in the ancient astronomy case, not only are the comparable objects different from Venus, but also they are different from each other, i.e.

$$\epsilon xBaRMx \neq \epsilon xBaSEx,$$

although the suggestion that these are mereological parts of Venus does seem very 'unconventional'.

But one unfortunate consequence of Edelberg's theory of such things, indeed any theory of counterparts surely, is that there cannot be any intersubjective means of referring to them, on that theory. For, on Edelberg's theory, all the phrase 'the Venus in my theory' could put into a reader's mind would be that reader's *counterpart* of the Venus in Edelberg's theory, not the Venus in Edelberg's theory itself. Contrariwise, however, the fact that Edelberg does successfully use an intersubjectively meaningful term means that the very same object can be on his mind as on his reader's: the very same object may be 'inside' a belief as 'outside' in the room. This shows that what is going on is simply that the same term — or one term and a pronominal replacement for it — arises at two *linguistic* places in some speech, and it has the same reference at each such *linguistic* place. Hence there is no essential referential difference between an anaphoric remark about someone's mind, like

Celia believed there was a woman in the room, but it was a man,

i.e.

$$Bc(\exists x)(Wx.Rx).M\epsilon xBc(Wx.Rx).$$

and the cross reference, for instance, in

There was a woman in the room. She was hungry.

i.e.

$$(\exists x)(Wx.Rx).H\epsilon x(Wx.Rx).$$

What has been the problem fundamentally, therefore, has been getting the cross-reference formalised first of all in this purely extensional kind of case. Yet this just requires extending the epsilon substitution method for existential statements, by means of a repetition of the appropriate epsilon term, as we saw before. The only difference in the intensional case is that, to obtain the required crossreferencing epsilon term one must first move from '$Bc(\exists x)(Wx.Rx)$' to '$(\exists x)Bc(Wx.Rx)$' via '$Bc(Wa.Ra)$' with $a = \epsilon x(Wx.Rx)$. The 'it' required is not 'that woman in the room', but 'what Celis believed was a woman in the room'.

It is now better understood how the epsilon calculus allows us to do this [Purdy, 1994; Egli and von Heusinger, 1995; Meyer Viol, 1995]. The starting point is the possibility illustrated in the theorem about Russellian definite descriptions

in Section 6, of separating out what otherwise, in the predicate calculus, would be a single sentence into a two-sentence piece of discourse, leaving the existence and uniqueness clauses in one place and putting the characterising remark in another. The point really starts to bite when there is no way to symbolise some anaphorically linked remarks in the predicate calculus, as in the above extensional case, where there is no uniqueness clause. This is what became a problem for the Fregean and Russellian logicians who woke up to the need to formalise anaphoric reference in the 1960s.

It can be seen in retrospect how it was lack of the epsilon calculus which was the major cause of the difficulty. Thus Geach, in an early discussion of the issue, went to the extremity of insisting that there could be no syllogism of the following form [Geach, 1962, p. 126]:

> A man has just drunk a pint of sulphuric acid.
>
> Nobody who drinks a pint of sulphuric acid lives through the day.
>
> So, he won't live through the day.

Instead, Geach said, there was only the existential conclusion:

> Some man who has just drunk a pint of sulphuric acid won't live through the day.

Certainly one can only conclude

$$(\exists x)(Mx.Dx.\neg Lx)$$

from

$$(\exists x)(Mx.Dx),$$

and

$$(x)(Dx \supset \neg Lx),$$

within Fregean predicate logic. But one can still conclude

$$\neg L\epsilon x(Mx.Dx),$$

within its extension: Hilbert's epsilon calculus.

And through inattention to that extension, Geach was entirely stumped later [Geach, 1967], when he discussed his famous example (3)

> Hob thinks a witch has blighted Bob's mare, and Nob wonders whether she (the same witch) killed Cob's sow,

i.e.

$$Th(\exists x)(Wx.Bxb).OnK\epsilon xTh(Wx.Bxb)c.$$

For he saw this could not be his (4)

$$(\exists x)(Wx.ThBxb.OnKxc),$$

or his (5)

$$(\exists x)(Th(Wx.Bxb).OnKxc).$$

But a reading of the second clause as

> Nob wonders whether the witch who blighted Bob's mare killed Cob's sow,

(c.f. Geach's 18) in which 'the witch who blighted Bob's mare killed Cob's sow' is analysed in the Russellian manner, as Geach's (20):

> just one witch blighted Bob's mare and she killed Cob's sow,

does not provide the required cross-reference — for one thing because of the uniqueness clause then involved. Of course the descriptive replacement for the personal pronoun 'she' in the Hilbertian expression, namely 'what Hob thinks is a witch that blighted Bob's mare' does not have any implication of uniqueness.

The inappropriateness of the uniqueness clause in Russellian analyses has been widely discussed. However it did not deter Neale, who much later wrote a whole book defending a largely Russellian account of definite descriptions, and cross-sentential anaphora. But he got no further with Geach's case above than proposing that 'she' might be localised to 'the witch we have been hearing about' [Neale, 1990, p. 221], thinking in general that definite descriptions merely should be relativised to the context. But a greater change is needed than that. For it is not that, in addition to any witches in the actual world, there are things in the actual world that are thought or believed to be witches. Geach's 'the same withch' is also inappropriate, on the same grounds. Thus the second clause in his (3) is not simply (as the present writer wrote in previous publication) '$OnK\epsilon x(Wx.Bxb)c$'.

But a large amount of the important, initial work in this area was done by another person very influenced by the Russellian tradition: Evans [Evans, 1977; Evans, 1980]. He also explicitly separated from Russell over the matter of uniqueness, for instance in connection with back-reference to a story about a man and a boy walking along a road one day [Evans, 1977, pp. 516–517]:

> One does not want to be committed, by this way of telling the story, to the existence of a day on which just one man and boy walked along a road. It was with this possibility in mind that I stated the requirement for the appropriate use of an *E*-type pronoun in terms of having answered, or *being prepared to answer upon demand*, the question 'He? Who?' or 'It? Which?'
>
> In order to effect this liberalisation we should allow the reference of the E-type pronoun to be fixed not only by predicative material explicitly

> in the antecedent clause, but also by material which the speaker supplies upon demand. This ruling has the effect of making the truth conditions of such remarks somewhat indeterminate; a determinate proposition will have been put forward only when the demand has been made and the material supplied.

It was Evans who popularised the name 'E-type pronoun' for the pronoun in such cases as

A Cambridge philosopher smoked a pipe, and he drank a lot of whisky,

i.e.

$$(\exists x)(Cx.Px).D\epsilon x(Cx.Px).$$

He also argued at length, in line with the above [Evans, 1977, p. 516], that what was distinctive about E-type pronouns was that such a conjunction of statements as this was not equivalent to

A Cambridge philosopher, who smoked a pipe, drank a lot of whisky,

i.e.

$$(\exists x)(Cx.Px.Dx).$$

Obviously the epsilon account supports this, since the contrast illustrates the point remarked before concerning cases which lack a uniqueness clause: only the expression which contains the relative pronoun can be symbolised in the predicate calculus, since to symbolise the personal pronoun its epsilon extension is needed.

Dynamic Predicate Logic agreed that some extension of Fregean predicate logic was needed, but took it that an existential quantifier could reach further than its scope, and in particular reach over further conjuncts, making, for instance,

$$((\exists x)Px).Qx,$$

still grammatical. However such a form was not equivalent to

$$Qx.(\exists x)Px,$$

[Groenendijk and Stokhof, 1991, p. 63], making dynamic conjunction non-commutative. Natural language conjunction, however, is still commutative, even when pronouns are involved. Thus we can say, for instance,

He may or may not be called Smith, but there is a murderer,

just as well as we can say

There is a murderer, although he may or may not be called Smith.

And extended forms of this kind of inversion occur in riddles.

The dynamic approach has difficulties in other areas, as well. Thus Evans is also noteworthy for realising [Evans, 1977, p. 517], that fixing the reference of an E-type pronoun can involve cancellation of explicit predicative material in the antecedent as in,

> A man jumped out of the crowd and fell in front of the horses. [Oh, no] he didn't jump, he was pushed.

Of course, the epsilon account has no difficulty with this feature, either, since this is

$$(\exists x)(Mx.Jx.Fx).\neg J\epsilon x(Mx.Jx.Fx).P\epsilon x(Mx.Jx.Fx).$$

Indeed, in some ways, the epsilon calculus' ability to formalise improper definite descriptions is the most central one. But the dynamic approach does not see the need for this facility. In fact, one theoretician in this area, van Eijck, has argued explicitly against an epsilon account, saying that this is unintuitive in the case of empty sets [van Eijck, 1993, p. 242], instead preferring an account which returns to Hilbert and Bernays' pre-suppositional iota and eta terms [van Eijck, 1993, p. 241]. Van Eijck forgets here the pragmatics of reference which Donnellan described, but the basic problem with his approach is that he is expecting a representation of, say,

A man is in the room,

just in terms of,

$$R\eta xMx,$$

which does not allow for the denial of the 'pre-supposition' as in Evans' case. The point confuses him elsewhere, as well, for to handle two entries of 'a man', as in

A man is in the room. A man is in the garden.

he thinks a dynamic logic is required, which employs different choice functions for 'a man', as the discourse proceeds. But there is no difficulty discriminating the two men by means of the two epsilon terms which naturally arise, in

$$(\exists x)(Mx.Rx).(\exists x)(Mx.Gx).$$

Gawron, Nerbonne and Peters moved some way in the direction of an epsilon account, by introducing choice functions to analyse E-type pronouns, but they also missed the crucial possibility of improper definite descriptions: in their case by wanting the clause containing the E-type pronoun to be invariably false when the antecedent is false [Gawron *et al.*, 1991, p. 357]. Of course it can have, in general, any truth value, since this is immaterial. Thus if the antecedent is false in Evans' story 'One day, a man and a boy were walking along a road, and the man said to the boy: "Would you like to be King?"' then the story is a fiction,

and one can continue it as one likes. The falsity of the antecedent ensures that the whole is entirely false, whatever truth value is attached to the remaining parts, i.e. whatever is further asserted. The fact that the epsilon terms for the man and the boy in question then refer to fictions, it should be noted, does not mean

$$\neg(\exists y)(y = \epsilon x(\exists z)(Mx.Bz.Wxz)),$$

or

$$\neg(\exists y)(y = \epsilon z(\exists x)(Mx.Bz.Wxz));$$

it merely means that

$$\neg(\exists x)(\exists z)(Mx.Bz.Wxz).$$

So the objects spoken of remain entirely extensional: it is only that their properties are mis-judged. Faust, for instance, was in fact a real person, even though the various stories about him were fictional; but the point has a much wider application (see [Slater, 1987; Slater, 1988a]).

Other grammarians have tried to handle this sort of issue in intensional contexts, but have returned to Husserlian 'intensional objects', or the 'counterparts' of actual individuals in doxastic alternative worlds. For example, Saarinen considers the following case drawn from Husserl [Saarinen, 1978, p. 277]:

> Bill believes that the lady on the stairs [is acquainted with] him, but John knows she is only a wax figure.

About this Saarinen says 'Both of the attitudes are of the wax lady, and yet all the relevant individuals in the doxastic worlds are not wax ladies but human beings' [Saarinen, 1978, p. 282]. However, as before, in addition to human beings in this world it is not that there are also human beings elsewhere — in people's mind — merely, also, that there are things in this world which are taken to be human beings, by people. Saarinen supports the opposite judgement with a Russellian, i.e. attributive reading of 'the lady on the stairs', but more important is his retention of the Meinongian idea that such an intensional object as the gold mountain has to be made of gold, and even that an impossible intensional object, like the round square, has to be both round and square. Thereby, of course, Saarinen misses the possibility that what Bill believes is the one and only lady on the stairs need not really be a lady on the stairs. But it is this possibility, exactly, which allows it to be just the plain, and everyday, physical wax figure which is the object of Bill's and John's attitudes. The form of Saarinen's case, if 'the lady on the stairs' is attributive, is

$$Bb(\exists x)((y)(Ly \equiv y = x).Abx).KjW\epsilon xBb((y)(Ly \equiv y = x).Abx),$$

but the second conjunct of this entails

$$\neg L\epsilon x((y)(Ly \equiv y = x).Abx),$$

since knowledge entails truth, and being a wax figure entails not being a lady.

As was mentioned before, it is contingent that the *lady on the stairs* is a lady on the stairs, but the source of this non-doxastic, and simply modal contingency cannot be properly seen until we link it with the linguistic possibility of improper, i.e. non-attributive definite descriptions. Seeing the source of this even more radical contingency is thus essentially linked to seeing how there can be *de re* attitudes. But it is also directly linked with the much more substantial programme of replacing such a metaphysical view as Husserl's simply with accurate linguistic analysis: 'philosophical problems arise through misconceptions of grammar' said Wittgenstein.

9 HIGHER-ORDER APPLICATIONS

Higher-order epsilon calculi have not been widely studied, although we met second-order epsilon terms briefly in Section 2, and epsilon numerical terms in Sections 5 and 6. Bernays' use of the epsilon operator in his formulation of Set Theory [Bernays, 1958] has been mentioned several times before — though he used sigma to denote it. The epsilon operator was also used in a formulation of type theory by Church [Church, 1940], and in the completeness proof built on that system by Henkin [Henkin, 1950] — though they used iota to denote it. The epsilon operator in intuitionistic type theories has been studied by others more recently [Bell, 1993b; DeVidi, 2004]. We can see the further need for higher-order calculi, now, in connection with a variety of other matters.

Not all anaphoric reference is first-order, for one thing, and epsilon terms of higher orders must be brought in to handle other cases. Thus we may have propositional, predicative, or numerical cases where the cross-reference is to something unspecific, as in

> John had a pleasant thought, and Mary had that thought too,
> John had a certain aloofness, and Mary had the same quality,
> John had a number of apples, and Mary had that number, also.

But use of epsilon calculi with propositional, predicative, or numerical variables is required, also, to solve a number of paradoxes which have arisen in these three areas. It is the possibility of indeterminacy in the reference of epsilon terms which does the trick.

In the first area, Williams is noteworthy for providing an iota term account of propositional referring phrases like 'what John thought', in extension of Prior's system of propositional quantification, which is usually associated with propositional operator theory [Williams, 1976]. But that requires, amongst other things, an implication of uniqueness which the corresponding epsilon term account does not have. Furthermore, Prior's propositional quantification is naturally taken to be substitutional, but there are some difficulties with this. In fact Loux [Loux, 1998, p. 150] has recently taken it to be final against the substitutional account

of propositional quantification that it cannot provide the truth conditions in such a case as:

$$(\exists p)(\text{it cannot be said in English that } p).$$

Certainly instantiation to an explicit English sentence is ruled out by the very nature of this case, but that does not prevent the formation of an appropriate propositional referring phrase, say 'that thought which cannot be said in English'. So, in a calculus in which the variables range over such referential phrases, a truth condition for

$$(\exists x)(x \text{ cannot be said in English}),$$

can be formulated in terms of the equivalent

$$\neg MSe\epsilon x \neg MSex.$$

An iota formalisation of 'the thought which cannot be said in English', somewhat in Williams' manner, would not be derivable, because of the lack of a uniqueness clause. But the epsilon definition of the now objectival quantifier, at this level, provides the appropriate truth condition while still allowing for the ineffability.

A comparable point comes in, also, in further explication of the propositional operator escape from propositional paradoxes such as the Liar, discussed before. For the propositional operator form 'it is true that p' may be formulated as a predication 'that p is true', and there are other truth predications of this kind besides those involving explicit propositional subject terms of the form 'that p'. Thus we can say 'what John thought is true', etc. The impossibility of self-reference with propositional operators then shows that there is not an identifiable proposition in connection with paradoxical sentences like

'what this sentence says is not true'.

For, if we could say with respect to this, something of the propositional operator form

what that sentence says is that p,

then we could get a contradiction. Hence what that sentence says cannot be made explicit. The sentence lacks a determinate sense, and the resulting ambiguity is the way the propositional operator approach therefore escapes the threat of paradox with 'semantic closure' [Slater, 1986b].

An epsilon analysis reveals this explicitly. Writing '$Sa|p$', now, for 'a says that p', with 'a' naming the sentence in question, and 'Sa' a predicate of propositional referential phrases (not exclusively of the form 'that p'), then we know that $Sa|\neg T\epsilon rSar$, but only if $(\exists !r)Sar$ does that lead to $\epsilon rSar = |\neg T\epsilon rSar$, from which to derive the contradiction $T\epsilon rSar \equiv \neg T\epsilon rSar$, using the propositional truth scheme: $T|p \equiv p$. It follow that $\epsilon rSar$, i.e. what a says, is indeterminate.

It was Sayward who first claimed that Tarski's sentential T-Schema presumed determinateness of sense [Sayward, 1987]. But a re-working of Goodstein's original propositional analysis of the Liar, which leads to a similar conclusion, is even more revealing. Thus if A says that everything that A says is false, we have

$$Sa|(y)(Say \supset \neg Ty),$$

and then, following Goodstein [Goodstein, 1958], we can prove that something A says is true, i.e.

$$(\exists x)(Sax.Tx),$$

and something A says is false, i.e.

$$(\exists x)(Sax.\neg Tx).$$

But with the variables here ranging over referring phrases to propositions rather than expressed sentences, as in Goodstein's and Sayward's presentations, what emerges further is that these two sayings need not be identifiable by means of expressed sentences. Indeed A may just utter one sentence, which must, as a result, carry both meanings. The result is that, while, in such a case, we might still talk about 'that sense of A's sentence in which it is true' ($\epsilon x(Sax.Tx)$) and 'that sense of A's sentence in which it is false' ($\epsilon x(Sax.\neg Tx)$) there need be no way to specify these senses, in explicit speech. Language must come to an end somewhere, in things which are, as a result, semantically ineffable, and we find that boundary here.

Thomason, in his argument against Fodor's Language of Thought ([Thomason, 1980], see also [Thomason, 1977]) and Asher and Kamp, in their argument against Montague's Intensional Logic ([Asher and Kamp, 1989, p. 94], see also [Asher and Kamp, 1986]), show that Fodor and Montague presume, by contrast, that a semantic specification of meaning is always available, since they take the expression relation between propositions and sentences to be 1-1. On this basis standard self-referential paradoxes with sentences can be mapped onto parallel paradoxes with propositions, making these writers' compositional theories of meaning inconsistent. But if the expression relation is not 1-1, as in Goodstein's case, then there is no problem, and in fact Goodstein proves his formalisation of indirect speech is consistent, at the end of his paper. Goodstein, however, forgot that the complements in propositional attitude constructions are not always phrases of the form 'that p': an epsilon account of general propositional referring phrases then allows them to have the appropriate indeterminacy.

It is not always noted how such indeterminacy enters into other paradoxes, for instance those involving predicates. In connection with the paradox of Heterologicality, for instance, overlooking indeterminacy has been explicit in at least one notable, and repeated place ([Copi, 1954, pp. 333–334], [Copi, 1967, pp. 346–347], [Copi, 1973, p. 301]), and that seems to have obscured the, otherwise obvious, exit to the paradox in question. For this paradox was introduced by Copi with

the definition "we shall use the word 'heterological' to designate the property possessed by words which designate properties not exemplified by themselves' . He then formalised this in an iota term manner as

$$Hs = (\exists F)(sDesF.(P)(sDesP \equiv P = F).\neg Fs),$$

and his proof of the contradiction goes in the following way. First, he shows that $H'H'$ entails $\neg H'H'$. That is quite straightforward, given 'H' Des H. Then he tries to show that $\neg H'H'$ entails $H'H'$. But there is a further assumption needed to obtain this: it is openly expressed in the next to the last line of Copi's proof, which receives the gloss from Copi "assuming 'H' is univocal'. Quite why Copi still repeatedly thought he had derived a contradiction which was inescapable is not too apparent, as a result, unless, by taking 'heterological' to 'designate the property...', he took its univocality to be somehow guaranteed by its iota-term definition. But he still needed a proof that there was one and only one such property. There is nothing which forces the designation of predicates to be 1-1, and in fact the denial of this is what must be concluded from the above proofs, by *Reductio*. For the pair of assumptions cannot both be true, so if anything designates the above H, it demonstrably does not do so uniquely — on pain of contradiction. The need for, and yet the unavailability of a proof of the existence and uniqueness of the property in question would have been more evident at the start if Copi had given the definition of Heterologicality in an *epsilon term* form:

$$Hs = \neg[\epsilon F(sDesF.(P)(sDesP \equiv P = F))]s.$$

For it is clear, in this explicit epsilon formulation of the property, that the property is not a constant, but a function of what it is a property of.

In other words, if we write s is heterological if and only if s is not self applicable, then we mask the proform 'self' which re-introduces 's' into the predicate. The constant expression 'is heterological' masks the variable in 'is not self-applicable'. The latter predicate is context-sensitive, because of the back-reference involved in the pronoun 'self'; as a result, 'is heterological' does not denote a constant property. The point clearly applies to the variable set associated with 'is not a member of itself', in Russell's Paradox, and the variable property associated with 'is a property which does not hold of itself' in the Paradox of Predication. The predicate 'is not a member of itself' involves a reflexive pronoun, which is a constant syntactic item, but in different applications it has a different referent. What is a set a member of iff it is not a member of itself? It is a member of its complement.

What misleads is the fact that in the lambda abstraction, and set abstraction expressions which are commonly taken to refer to the associated properties and sets, all the variables are bound. Aren't the properties and sets involved, therefore, necessarily non-context-sensitive? But while we might try to represent Heterologicality as $\lambda x(x$ does not possess $\epsilon P(x$ stands for $P))$, this abstract is drawn from a whole sentence, and it is only a part of that sentence, namely 'does not possess

the property it (sic) stands for', which expresses the property involved, and which relates, therefore, to any associated set. If $x \notin x$, that is not a property of x but a truth about it, it must be remembered; and the property it then possesses is, in fact, not only context-sensitive, but also ambiguous, since there are two ways that simply 'x' can be its subject. Taking the first 'x' as subject, the related property 'being a member of its complement' is '$\lambda y(y \notin x)$' in this context, while taking the second 'x' as subject, the related property 'being non-self-membered' is '$\lambda y(x \notin y)$', in this context. The context sensitivity arises because words such as 'its' and 'self' have a variable reference, being pro-forms merely standing in place of a repeat of the subject. What was missing in Fregean abstracts was, therefore, first of all, the identification a single, *subject* place where the variable is bound. To get a property which is not context sensitive, repeated variables must be represented by means of ordered sets. Thus Rxx says that the ordered pair x, x belongs to the set of ordered pairs y, z such that Ryz, and then with 'Ryz' as '$y \notin z$' there is no contradiction.

A more general point about grammar, in this area, needs to be noted before the corresponding resolution of the Paradox is Predication is clearly obtained. For grammar requires that the predicate 'is a P' expresses a concept, while it is the nominal phrase 'being a P' which refers to it [Kneale and Kneale, 1962, 585–587]. This is not a distinction customarily symbolised in second-order logic, which is to that logic's great cost, since, amongst other things, there is no way to question whether what is expressed by some predicate is a single concept. Thus, if we write

$$(\exists F)(F = x.\neg Fx) \equiv Gx,$$

then the double duty which terms like 'F' are doing — to express a concept, as in 'Fx', and to refer to a concept, as in '$F = x$' — makes it seem automatic, by mere abbreviation, that there is a single concept, G, which the predicate on the left is expressing, namely

$$\lambda x(\exists F)(F = x.\neg Fx).$$

But it is this which leads to the Paradox of Predication, upon substituting 'G' for 'x'. Let us therefore write 'x has F' in place of 'Fx', enabling 'F' throughout to be a referring term, then there is nothing to guarantee that

$$(\exists G)(x)((\exists F)(F = x.\neg(x \text{ has } F)) \equiv x \text{ has } G).$$

Maybe '$\lambda x(\exists F)(F = x.\neg(x\ hasF))H$' is still equivalent to '$(\exists F)(F = H.\neg(H$ has $F))$', but there is nothing to show there is a concept G for which this is equivalent to 'H has G'. Indeed, if there was then there would be a contradiction. So there isn't, and the epsilon term

$$\epsilon G(x)((\exists F)(F = x.\neg(x \text{ has } F)) \equiv x \text{ has } G),$$

as a consequence, must have an indeterminate reference.

How is it that '$(\exists F)(F = a.\neg(a$ has $F))$', and '$(\exists F)(F = b.\neg(b$ has $F))$', do not say a and b have the same property? The common predicate of a and b is 'is a

property it does not possess' and this contains an item with a variable reference, namely 'it', like the 'itself', 'its', and 'self' before.

But the second-order epsilon calculus which might clarify these matters is even further from full formulation. Gamut, for instance, had the following to say about the semantics of standard second-order logic [Gamut, 1991, pp. 170–171]:

> ... how is quantification over properties to be interpreted? Given that properties are to be interpreted as subsets of the domain, and that the quantifiers are supposed to apply to all properties, presumably quantifiers must apply to *all subsets of the domain* D...So now an assignment will assign elements of the domain D to individual variables, and subsets of D,...to predicate variables. And how should second-order predicate constants be interpreted? They express properties of, or relations between, properties of entities. So just as a unary first-order predicate is interpreted as a set of entities, a unary second-order predicate is interpreted as a set of sets of entities...$I(C)$, the interpretation of the second-order predicate constant *is a color*, is then $\{I(R), I(G), ...\}$, in which R is the first-order predicate constant *is red*, G is *is green*, and so on.

The problem with this, and the largely similar Henkin semantics, is that it simply ramifies the problem with the standard set semantics of predicate logic. For, since there are predicates which do not determine a unit, i.e. whose instances are not discrete, then for some predicates there is no determinate number, and so no set associated with them. In fact both redness and colour are of this kind, since red things may overlap, or even be entirely included one within another, and it is arbitrary where one colour starts and another begins. One needs, for a start, a symbolism which allows some predicates to have no determinate number of instances — which the epsilon, intensional treatment of properties respects.

Indeterminacy of number enters into these matters in a variety of other ways, as well. Cantor, for instance, realised that there could not be a determinate number of all things because of his, Cantor's Paradox, so the possibility of innumerability, in principle, is a part of his Set Theory. Indeed, the crucial question, for Cantor, of the 'limitation of size' was consequently a preoccupation of many contemporary thinkers, wondering which infinities might still be consistently numbered. Cantor could, in the end, only offer as a postulate that Aleph Zero is consistent [Hallett, 1984, p. 175], and it is necessary, of course, to incorporate the same assumption, as an Axiom of Infinity, in all infinitary Set Theories. But a stronger point, more in line with Hilbert's finitism, was made by Tiles [Tiles, 1989, p. 63], who reminds us that there is, properly, an inconsistency even in Aleph Zero once we attend not to the 'cardinality' or 'power' of the natural numbers, but to the size or number of them. For, we may argue, if there were a determinate number of natural numbers it would be the same as the number of the even numbers, because of the 1-1 correspondence, but also it would be bigger, since the even numbers are a proper

part of the natural numbers. Hence there is no determinate *number* of the natural numbers, even though they have a determinate power (c.f. [Boolos, 1998, p. 306]).

The question of the number of the natural numbers was historically seen as Galileo's Paradox: it teased most number theorists until the end of the nineteenth century, even if it has been forgotten by succeeding generations of set theoreticians. But its resolution, in the above way, is in a similar manner to the resolution of Cantor's Paradox: a presupposition of the determinacy of the number of some discrete elements must be negated. Tiles thinks that Cantor's proofs of non-denumerability, nevertheless, still require us to discriminate between infinities in terms of their size [Tiles, 1989, p. 63]. But, as Crispin Wright has pointed out, the diagonal argument merely shows that we need to discriminate amongst the denumerable between what can be determinately enumerated, and what cannot [Wright, 1985, p. 134]. Certainly the natural numbers and the even numbers have the same power, since they can be put into 1-1 correspondence; but the natural numbers and the ordinal numbers of those Turing Machines which halt, amongst the totality of Turing Machines in some ordered, denumerable sequence, cannot be put into 1-1 correspondence, and so they have different powers while both are denumerable. As Borel also saw, in connection with Richard's Paradox [Martin-Löf, 1970, p. 44], in certain cases we are simply unable to specify an enumerating function for the things in question, even though they are definitely denumerable. Wittgenstein also realised that no difference in size is involved in connection with diagonalisation [Rodych, 2000]. Diagonalisation merely allows us to generate a further denumerable sequence given we can enumerate some others.

But indeterminacy of number, in yet another sense, is crucial to the resolution of one final paradox about numbers: Berry's Paradox. Surely the least ordinal not definable in less than 19 words is definable in less than 19 words, since 'the least ordinal not definable in less than 19 words' has less than 19 words in it? In Priest's recent treatment [Priest, 1994, p. 29], there occurs the line

$$lon(x) \notin x,$$

which is meant to formalise the supposedly evident truth that the least ordinal not in a set is not in that set. Reading 'lon' as 'the least ordinal not in' Priest wants to say that

$$lon(x) \notin x$$

holds for all sets x, and so, in particular, it holds for

$$lon(DN19) \notin DN19,$$

where $DN19$ is the set of finite ordinals definable in less than 19 words. But he also wants to say

$$lon(DN19) \in DN19,$$

since 'the least ordinal not in the set of finite ordinals definable in less than 19 words' defines a finite ordinal in less than 19 words — because of its length.

Now certainly *any* least ordinal not in a set is not in that set, i.e. it is necessary that

$$(x)(y)([Oy.y \notin x.(z)((Oz.z \notin x) \supset y \leq z)] \supset y \notin x).$$

But that does not employ a referential term 'the least ordinal'. So if we want to make a referential remark of the kind Priest had in mind, we must say, instead

$$(x)(\epsilon y[Oy.y \notin x.(z)((Oz.z \notin x) \supset y \leq z)] \notin x),$$

but that is then no longer a universal truth, and so its instantiation to $DN19$ does not follow. What would be a universal truth, following from the previous one, would be

$$(x)([Oa.a \notin x.(z)((Oz.z \notin x) \supset a \leq z)] \supset \epsilon y[Oy.y \notin x.(z)((Oz.z \notin x) \supset y \leq z)] \notin x),$$

where

$$a = \epsilon y[Oy.y \notin x.(z)((Oz.z \notin x) \supset y \leq z)].$$

But there is now no problem with

$$lon(DN19) \notin DN19,$$

since there is now no requirement to also affirm the reverse. For the denial of the antecedent of this last conditional, in specific cases, is quite possible, allowing a way out from Berry's Paradox through the associated epsilon term being non-attributive, and so not having a determinate referent. Every ordinal is then 'definable' using the associated epsilon term in less than 19 words, and so there is no determinate least ordinal not in the set so 'definable'. The phrase then operates non-semantically like a demonstrative such as 'this'.

Priest may have thought to avoid such demonstrative reference, since he says [Priest, 1994, p. 28] that 'something is definable iff there is a (non-indexical) noun phrase that refers to it'. But the trouble is that that does not guarantee that the referring expression 'the least ordinal not in the set of those definable in less than 19 words' functions attributively, and so enables one to locate its referent through its sense. In an earlier treatment of this same paradox Priest, like Copi, Thomason, and Asher and Kamp, derived a contradiction assuming that the denotation relation was 1-1 [Preist, 1983, p. 162]. But again the conclusion must be that all that Priest has proved is that if the denotation relation is univocal then there is a contradiction. Hence, we may deduce, the denotation relation is not univocal — although we must also realise that that does not mean a more liberal semantic denotation relation could be defined. In brief, if 'the least ordinal not definable in less than 19 words' defined, i.e. uniquely referred to some object, then that object both would and would not be definable in less than 19 words. Hence, directly against Priest's pre-supposition, some identifying descriptions cannot have a unique denotation. But that does not mean that 'the least ordinal not definable in less than 19 words' is, for instance, semantically ambiguous. To think so is to

forget the function of demonstratives like 'this', and purely referential terms like Donnellan's 'the man drinking Martini', whose semantics does not determine, not just a unique referent, but any referent at all.

For what these cases show instead is simply that there is something beyond semantics: pragmatics. In fact 'the least ordinal not definable in less than 19 words' is *incapable* of describing what it refers to, and so what it refers to has to be semantically ineffable. But that merely means that the phrase is non-attributive, and gains its reference pragmatically from the context in which it is used, just like 'the man drinking Martini'. It is in such pragmatic cases that we find the paradigms for where language comes to an end, and yet meaning is still expressed — for instance, simply by means of the gesture which accompanies 'this', or some other purely referential expression. Epsilon terms are the formal representation of such demonstratives, as we have now seen in a very wide variety of cases.

BIBLIOGRAPHY

[Ackermann, 1924] W. Ackermann. Begründung des tertium non datur mittels der hilbertischen theorie der widerspruchsfreiheit. *Mathematische Annalen*, 93:1–36, 1924.

[Ackermann, 1937–8] W. Ackermann. Mengentheoretische begründung der logik. *Mathematische Annalen*, 115:1–22, 1937–8.

[Ackermann, 1940] W. Ackermann. Zur widerspruchsfreiheit der zahlentheorie. *Mathematische Annalen*, 117:162–194, 1940.

[Asher and Kamp, 1986] N. Asher and H. Kamp. The knower's paradox and representational theories of the attitudes. In J. Halpern, editor, *Theoretical Aspects of Reasoning about Knowledge*. Morgan Kaufmann, San Mateo CA, 1986.

[Asher and Kamp, 1989] N. Asher and H. Kamp. Self-reference, attitudes and paradox. In G. Chierchia, B. H. Partee, and R. Turner, editors, *Properties, Types and Meaning*. Kluwer, Dordrecht, 1989.

[Asser, 1957] G. Asser. Theorie der logischen auswahlfunktionen. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 3:30–68, 1957.

[Bell, 1993a] J. L. Bell. Hilbert's ϵ-operator and classical logic. *Journal of Philosophical Logic*, 22:1–18, 1993.

[Bell, 1993b] J. L. Bell. Hilbert's ϵ-operator in intuitionistic type theories. *Mathematical Logic Quarterly*, 39:323–337, 1993.

[Benacerraf, 1965] P. Benacerraf. What numbers could not be. *Philosophical Review*, 75:47–73, 1965.

[Bernays, 1958] P. Bernays. *Axiomatic Set Theory*. North Holland, Dordrecht, 1958.

[Bertolet, 1980] R. Bertolet. The semantic significance of donnellan's distinction. *Philosophical Studies*, 37:281–288, 1980.

[Boolos, 1975] G. Boolos. On second-order logic. *The Journal of Philosophy*, 72:509–527, 1975.

[Boolos, 1984] G. Boolos. To be is to be a value of a variable (or to be some values of some variables). *The Journal of Philosophy*, 81:430–449, 1984.

[Boolos, 1998] G. Boolos. *Logic, Logic, and Logic*. Harvard University Press, Cambridge MA, 1998.

[Bostock, 1974] D. Bostock. *Logic and Arithmetic*, volume I. O.U.P., Oxford, 1974.

[Bostock, 1979] D. Bostock. *Logic and Arithmetic*, volume II. O.U.P., Oxford, 1979.

[Bourbaki, 1954] N. Bourbaki. *Éléments de Mathématique*. Hermann, Paris, 1954.

[Bunt, 1985] H. C. Bunt. *Mass Terms and Model-Theoretic Semantics*. C.U.P., Cambridge, 1985.

[Church, 1940] A. Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5:56–68, 1940.

[Cocchiarella, 1988] N. B. Cocchiarella. Predication versus membership in the distinction between logic as language and logic as calculus. *Synthese*, 77:37–72, 1988.

[Copi, 1954] I. Copi. *Symbolic Logic*. Macmillan, New York, 2 edition, 1954.
[Copi, 1967] I. Copi. *Symbolic Logic*. Macmillan, New York, 3 edition, 1967.
[Copi, 1973] I. Copi. *Symbolic Logic*. Macmillan, New York, 4 edition, 1973.
[Curry, 1963] H. B. Curry. *Foundations of Mathematical Logic*. McGraw-Hill, New York, 1963.
[da Costa, 1980] N. C. A. da Costa. A model-theoretical approach to variable binding term operators. In A. I. Arruda et al., editors, *Mathematical Logic in Latin America*. North Holland, Dordrecht, 1980.
[Detlefsen, 1986] M. Detlefsen. *Hilbert's Program*. Reidel, Dordrecht, 1986.
[DeVidi, 1995] D. DeVidi. Intuitionistic ϵ- and τ- calculi. *Mathematical Logic Quarterly*, 41:523–546, 1995.
[DeVidi, 2004] D. DeVidi. Choice principles and constructive logics. *Philosophia Mathematica*, 12:222–243, 2004.
[Devitt, 1974] M. Devitt. Singular terms. *The Journal of Philosophy*, 71:183–205, 1974.
[Donnellan, 1966] K. Donnellan. Reference and definite descriptions. *Philosophical Review*, 75:281–304, 1966.
[Dragalin, 1974] A. C. Dragalin. Intuitionistic logic and hilbert's ϵ-symbol. In *History and Methodology of Sciences, XVI: Mathematics, Mechanics*. MGU, Moscow, 1974.
[Dunn and Belnap, 1968] M. Dunn and N. D. Belnap. The substitution interpretation of the quantifiers. *Nous*, 2:177–186, 1968.
[Edelberg, 1986] W. Edelberg. A new puzzle about intentional identity. *Journal of Philosophical Logic*, 15:1–25, 1986.
[Edelberg, 1995] W. Edelberg. A perspectivalist semantics for the attitudes. *Nous*, 29:316–342, 1995.
[Egli and von Heusinger, 1995] U. Egli and K. von Heusinger. The ϵ operator and e-type pronouns. In U. Egli et al., editors, *Lexical Knowledge in the Organisation of Language*. Benjamins, Amsterdam, 1995.
[Evans, 1977] G. Evans. Pronouns, quantifiers and relative clauses. *Canadian Journal of Philosophy*, 7:467–536, 1977.
[Evans, 1980] G. Evans. Pronouns. *Linguistic Inquiry*, 11:337–362, 1980.
[Ferrari, 1987] P. L. Ferrari. A note on a proof of hilbert's 2nd ϵ-theorem. *Journal of Symbolic Logic*, 52:214–215, 1987.
[Ferrari, 1989] P. L. Ferrari. The rank function and hilbert's second ϵ-theorem. *Zeitschrift für Mathematische Logik*, 35:367–373, 1989.
[Fine, 1985] K. Fine. *Reasoning with Arbitrary Objects*. Blackwell, Oxford, 1985.
[Fitting, 1972] M. Fitting. An epsilon calculus system for first order s4. In *Conference on Mathematical Logic, Bedford College 1970*. Springer Verlag, Munich, 1972.
[Fitting, 1975] M. Fitting. A modal logic epsilon calculus. *Notre Dame Journal of Formal Logic*, 16:1–16, 1975.
[Fitting and Mendelsohn, 1998] M. Fitting and R. L. Mendelsohn. *First-Order Modal Logic*. Kluwer, Dordrecht, 1998.
[Flannagan, 1975] T. B. Flannagan. On an extension of hilbert's ii ϵ-theorem. *Journal of Symbolic Logic*, 40:393–397, 1975.
[Fraenkel and Bar-Hillel, 1958] A. A. Fraenkel and Y. Bar-Hillel. *Foundations of Set Theory*. North Holland, Amsterdam, 1958.
[Frege, 1971] G. Frege. Über die grundlagen der geometry. In E. H. Kluge, editor, *On the Foundations of Geometry and Formal Theories of Arithmetic*. Yale University Press, New Haven, 1971.
[Gamut, 1991] L. T. F. Gamut. *Logic, Language and Meaning*, volume I. University of Chicago Press, Chicago, 1991.
[Gawron *et al.*, 1991] J. M. Gawron, J. Nerbonne, and S. Peters. The absorption principle and e-type anaphora. In J. Barwise et al., editors, *Situation Theory and its Applications*, volume II. CSLI, Stanford, 1991.
[Geach, 1962] P. T. Geach. *Reference and Generality*. Cornell University Press, Ithaca, 1962.
[Geach, 1967] P. T. Geach. Intentional identity. *The Journal of Philosophy*, 64:627–632, 1967.
[Gentzen, 1934–5] G. Gentzen. Untersuchungen über das logischen schliessen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1934-5.
[Goddard and Routley, 1973] L. Goddard and R. Routley. *The Logic of Significance and Context*. Scottish Academic Press, Aberdeen, 1973.

[Gödel, 1969] K. Gödel. An interpretation of the intuitionistic sentential calculus. In J. Hintikka, editor, *The Philosophy of Mathematics*. O.U.P., Oxford, 1969.
[Gödel, 1980] K. Gödel. On a hitherto unexploited extension of the finitary standpoint. *Journal of Philosophical Logic*, 9:133–142, 1980.
[Goodman and Myhill, 1978] N. Goodman and J. Myhill. Choice implies excluded middle. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 24:461, 1978.
[Goodstein, 1958] R. L. Goodstein. On the formalisation of indirect discourse. *Journal of Symbolic Logic*, 23:417–419, 1958.
[Groenendijk and Stokhof, 1991] J. Groenendijk and M. Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100, 1991.
[Hallett, 1984] M. Hallett. *Cantorian Set Theory and Limitation of Size*. Clarendon, Oxford, 1984.
[Hazen, 1987] A. Hazen. Natural deduction and hilbert's ϵ-operator. *Journal of Philosophical Logic*, 16:411–421, 1987.
[Henkin, 1950] L. Henkin. Completeness in the theory of types. *Journal of Symbolic Logic*, 15:81–91, 1950.
[Hermes, 1965] H. Hermes. *Eine Termlogik mit Auswahloperator*. Springer Verlag, Berlin, 1965.
[Hilbert, 1923] D. Hilbert. Die logischen grundlagen der mathematik. *Mathematische Annalen*, 88:151–165, 1923.
[Hilbert, 1967] D. Hilbert. On the infinite. In J. van Heijenhoort, editor, *From Frege to Gödel*. Harvard University Press, Cambridge MA, 1967.
[Hilbert and Bernays, 1970] D. Hilbert and P. Bernays. *Grundlagen der Mathematik*. Springer, Berlin, 2 edition, 1970.
[Hughes and Cresswell, 1968] G. E. Hughes and M. J. Cresswell. *An Introduction to Modal Logic*. Methuen, London, 1968.
[Hughes and Cresswell, 1996] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. Routledge, London, 1996.
[Jeffrey, 1967] R. Jeffrey. *Formal Logic: Its Scope and Limits*. McGraw-Hill, New York, 1 edition, 1967.
[Kalish and Montague, 1964] D. Kalish and R. Montague. *Logic; Techniques of Formal Reasoning*. Harcourt, Brace and World, Inc, New York, 1964.
[Kneale and Kneale, 1962] W. Kneale and M. Kneale. *The Development of Logic*. Clarendon Press, Oxford, 1962.
[Kneebone, 1963] G. T. Kneebone. *Mathematical Logic and the Foundations of Mathematics*. Van Nostrand, Dordrecht, 1963.
[Kreisel, 1951] G. Kreisel. On the interpretation of non-finitist proofs i. *Journal of Symbolic Logic*, 16:241–267, 1951.
[Kreisel, 1952] G. Kreisel. On the interpretation of non-finitist proofs ii. *Journal of Symbolic Logic*, 17:43–58, 1952.
[Kreisel, 1964] G. Kreisel. Hilbert's programme. In P. Benacerraf and H. Putnam, editors, *Philosophy of Mathematics*. Prentice-Hall, Englewood Cliffs, 1964.
[Leisenring, 1969] A. C. Leisenring. *Mathematical Logic and Hilbert's ϵ-symbol*. Macdonald, London, 1969.
[Leivant, 1973] D. Leivant. *Existential Instantiation in a System of Natural Deduction for Intuitionistic Arithmetic*. Stichting Mathematical Centrum, Amsterdam, 1973.
[Loux, 1998] M. J. Loux. *Metaphysics; a Contemporary Introduction*. Routledge, London, 1998.
[Maehara, 1955] S. Maehara. The predicate calculus with ϵ-symbol. *Journal of the Mathematical Society of Japan*, 7:323–344, 1955.
[Marciszewski, 1981] W. Marciszewski. *Dictionary of Logic*. Martinus Nijhoff, The Hague, 1981.
[Martin-Löf, 1970] P. Martin-Löf. *Notes on Constructive Mathematics*. Almqvist and Wiksell, Stockholm, 1970.
[Meyer Viol, 1995] W. P. M. Meyer Viol. *Instantial Logic*. ILLC, Amsterdam, 1995.
[Mints, 1974] G. E. Mints. Heyting's predicate calculus with the epsilon symbol. In *Notes of Scientific Seminars, T40: Research in Constructive Mathematics, and Mathematical Logic*, volume 6. LOMI, Leningrad, 1974.
[Mints, 1977] G. E. Mints. Heyting predicate calculus with epsilon symbol. *Journal of Soviet Mathematics*, 8:317–323, 1977.
[Mints, 1982] G. E. Mints. Simplified consistency proof for arithmetic. *Proceedings of the Estonian Academy of Sciences, Fiz-Math*, 31:376–382, 1982.

[Mints, 1989] G. E. Mints. Epsilon substitution method for the theory of hereditarily finite sets. *Proceedings of the Estonian Academy of Sciences, Fiz-Math*, 38:154–164, 1989.
[Mints, 1994] G. E. Mints. Gentzen-type systems and hilbert's epsilon substitution method, i. In D. Prawitz, B. Skyrms, and D. Westerstahl, editors, *Logic, Methodology and Philosophy of Science*. Elsevier Science, Amsterdam, ix edition, 1994.
[Mints, 1996] G. E. Mints. Strong termination for the epsilon substitution method. *Journal of Symbolic Logic*, 61:1193–1205, 1996.
[Mints, 1997] G. E. Mints. A method of epsilon substitution for predicate logic with equality. *Journal of Mathematical Sciences*, 87:3234–3252, 1997.
[Mints and Tupailo, 1999] G. E. Mints and S. Tupailo. Epsilon substitution method for ramified languages and delta-1-1 comprehension axiom. In A Cantini et al., editors, *Logic and Foundations of Mathematics*. Kluwer, Dordrecht, 1999.
[Mints *et al.*, 1996] G. E. Mints, S. Tupailo, and W. Buchholz. Epsilon substitution method for elementary analysis. *Archive for Mathematical Logic*, 35:103–130, 1996.
[Montague, 1963] R. Montague. Syntactical treatments of modality, with corollaries on reflection principles and finite axiomatisability. *Acta Philosophica Fennica*, 16:155–167, 1963.
[Moore, 1982] G. H. Moore. *Zermelo's Axiom of Choice*. Springer-Verlag, New York, 1982.
[Neale, 1990] S. Neale. *Descriptions*. M.I.T. Press, Cambridge MA, 1990.
[Priest, 1979] G. G. Priest. Indefinite descriptions. *Logique et Analyse*, 22:5–21, 1979.
[Preist, 1983] G. G. Priest. The logical paradoxes and the law of the excluded middle. *Philosophical Quarterly*, 33:160–165, 1983.
[Priest, 1984] G. G. Priest. Semantic closure. *Studia Logica*, XLIII 1/2:117–129, 1984.
[Priest, 1994] G. G. Priest. The structure of the paradoxes of self-reference. *Mind*, 103:25–34, 1994.
[Prior, 1971] A. N. Prior. *Objects of Thought*. O.U.P, Oxford, 1971.
[Purdy, 1994] W. C. Purdy. A variable-free logic for anaphora. In P. Humphreys, editor, *Patrick Suppes: Scientific Philosopher*, volume 3. Kluwer, Dordrecht, 1994.
[Quine, 1960] W. V. O. Quine. *Word and Object*. Wiley, New York, 1960.
[Rasiowa, 1956] H. Rasiowa. On the ϵ-theorems. *Fundamenta Mathematicae*, 43:156–165, 1956.
[Rodych, 2000] V. Rodych. Wittgenstein's critique of set theory. *The Southern Journal of Philosophy*, 38:281–319, 2000.
[Rosser, 1953] J. B. Rosser. *Logic for Mathematicians*. McGraw-Hill, New York, 1953.
[Routley, 1969] R. Routley. A simple natural deduction system. *Logique et Analyse*, 12:129–152, 1969.
[Routley, 1977] R. Routley. Choice and descriptions in enriched intensional languages 2, and 3. In E. Morscher, J. Czermak, and P. Weingartner, editors, *Problems in Logic and Ontology*. Akademische Druck und Velagsanstalt, Graz, 1977.
[Routley, 1980] R. Routley. *Exploring Meinong's Jungle*. Philosophy Department, R.S.S.S., A.N.U., Canberra, 1980.
[Routley *et al.*, 1974] R. Routley, R. Meyer, and L. Goddard. Choice and descriptions in enriched intensional languages 1. *Journal of Philosophical Logic*, 3:291–316, 1974.
[Saarinen, 1978] E. Saarinen. Intentional identity interpreted: A case study of the relations among quantifiers, pronouns, and propositional attitudes. In E Saarinen, editor, *Game Theoretical Semantics*. Reidel, Dordrecht, 1978.
[Sayward, 1987] C. Sayward. Prior's theory of truth. *Analysis*, 47:83–87, 1987.
[Slater, 1986a] B. H. Slater. E-type pronouns and ϵ-terms. *Canadian Journal of Philosophy*, 16:27–38, 1986.
[Slater, 1986b] B. H. Slater. Prior's analytic. *Analysis*, 46:76–81, 1986.
[Slater, 1987] B. H. Slater. Fictions. *British Journal of Aesthetics*, 27:145–155, 1987.
[Slater, 1988a] B. H. Slater. Excluding the middle. *Critica*, XX:55–71, 1988.
[Slater, 1988b] B. H. Slater. Hilbertian reference. *Nous*, 22:283–297, 1988.
[Slater, 1988c] B. H. Slater. Intensional identities. *Logique et Analyse*, 121–2:93–107, 1988.
[Slater, 1989] B. H. Slater. Modal semantics. *Logique et Analyse*, 127–8:195–209, 1989.
[Slater, 1990] B. H. Slater. Using hilbert's calculus. *Logique et Analyse*, 129–130:45–67, 1990.
[Slater, 1992a] B. H. Slater. Review of m. detlefsen's hilbert's program (reidel, dordrecht 1986). *Nous*, 26:513–514, 1992.
[Slater, 1992b] B. H. Slater. Routley's formulation of transparency. *History and Philosophy of Logic*, 13:215–224, 1992.

[Slater, 1992c] B. H. Slater. Self-reference with operators. *Logique et Analyse*, 139–140:335–340, 1992.
[Slater, 1994a] B. H. Slater. The epsilon calculus' problematic. *Philosophical Papers*, XXIII:217–242, 1994.
[Slater, 1994b] B. H. Slater. *Intensional Logic*. Avebury, Aldershot, 1994.
[Smirnov, 1971] V. A. Smirnov. Elimination des terms ϵ dans la logique intuitioniste. *Revue Internationale de Philosophie*, 98:512–519, 1971.
[Smirnov, 1978] V. A. Smirnov. Theory of quantification and ϵ-calculi. In J. Hintikka, I. Niiniluoto, and E. Saarinen, editors, *Essays on Mathematical and Philosophical Logic*. Reidel, Dordrecht, 1978.
[Stalnaker and Thomason, 1968] R. C. Stalnaker and R. H. Thomason. Abstraction in first-order modal logic. *Theoria*, 3:203–207, 1968.
[Steen, 1972] S. W. P. Steen. *Mathematical Logic*. C.U.P., Cambridge, 1972.
[Tait, 1965] W. Tait. The substitution method. *Journal of Symbolic Logic*, 30:175–192, 1965.
[Tait, 1981] W. Tait. Finitism. *Journal of Philosophy*, 78:524–546, 1981.
[Thomason, 1977] R. Thomason. Indirect discourse is not quotational. *Monist*, 60:340–354, 1977.
[Thomason, 1980] R. Thomason. A note on syntactical treatments of modality. *Synthese*, 44:391–395, 1980.
[Thomason and Stalnaker, 1968] R. Thomason and R. C. Stalnaker. Modality and reference. *Nous*, 2:359–372, 1968.
[Tiles, 1989] M. Tiles. *The Philosophy of Set Theory*. Blackwell, Oxford, 1989.
[Tiles, 1991] M. Tiles. *Mathematics and the Image of Reason*. Routledge, London, 1991.
[van Eijck, 1993] J. van Eijck. The dynamics of description. *Journal of Semantics*, 10:239–267, 1993.
[von Heusinger and Egli, 2000] K. von Heusinger and U. Egli, editors. *Reference and Anaphoric Relations*. Kluwer, Dordrecht, 2000.
[von Neumann, 1927] J. von Neumann. Zur hilbertishen beweistheorie. *Mathematische Zeitschrift*, 26:1–46, 1927.
[Wang, 1962] H. Wang. *A Survey of Mathematical Logic*. Science Press, Peking, 1962.
[Wessels, 1977] L. Wessels. Cut elimination in a gentzen-style ϵ-calculus without identity. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 23:527–538, 1977.
[Wessels, 1978] L. Wessels. A cut-free sequent calculus for hilbert's ϵ-symbol. *Journal of Symbolic Logic*, 43:525–536, 1978.
[Weyl, 1944] H. Weyl. David hilbert and his mathematical work. *Bulletin of the American Mathematical Society*, 50:612–654, 1944.
[Williams, 1976] C. J. F. Williams. *What is Truth?* C.U.P., Cambridge, 1976.
[Wright, 1983] C. Wright. *Frege's Conception of Numbers as Objects*. Aberdeen University Press, Aberdeen, 1983.
[Wright, 1985] C. Wright. Skolem and the skeptic. *Proceedings of the Aristotelian Society*, 59:117–37, 1985.
[Yashuhara, 1982] M. Yasuhara. Cut elimination in ϵ-calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 28:311–316, 1982.
[Zach, 2003] R. Zach. The practice of finitism: Epsilon calculus and consistency proofs in hilbert's program. *Synthese*, 137:211–259, 2003.

GÖDEL'S LOGIC

Mark van Atten and Juliette Kennedy

The emphasis will be on Gödel's concrete theorems in mathematical logic. For his work in set theory, we refer to Volume 6 of this *Handbook*. The first part of the present chapter, on Gödel's work in classical logic, was written by Kennedy, the second part, on Gödel's work in intuitionistic logic, by van Atten. Responsibility for the chapter as a whole is accepted by both, as the two parts have been subjected to mutual discussion, criticism, and emendations. In the references to Gödel's own papers, we have followed the usage in the Collected Works.

Gödel first published a result in 1930 (his completeness theorem), and did so for the last time in 1958 (the Dialectica Interpretation); the latter was largely based on ideas he had already presented in a lecture seventeen years earlier [Gödel, *1941].

1 GÖDEL'S RESULTS IN CLASSICAL LOGIC AND ARITHMETIC

1.1 Introduction

In this part of our chapter we consider Gödel's principal theorems in classical logic. The section is not comprehensive; not treated here is Gödel's work on cases of the decision problem for predicate logic,[1] as well as the results for propositional logic 1932c, 1933, 1933a.[2] Also not treated are the myriad of results and observations contained in Gödel's reviews. Finally, we also fail to include a discussion of Gödel's 'Ontologische Beweis.' Though it is, at first glance, an exercise in modal logic, in the authors' opinion it is more properly treated in a discussion of Gödel's philosophical work (although the distinction in Gödel's work between the philosophical and the technical must nearly always remain an artificial one.)

1.2 The Completeness Theorem

Introduction

The completeness question for the first order predicate calculus was stated precisely and in print for the first time in 1928 by Hilbert and Ackermann in their text *Grundzüge der theoretischen Logik*.[3] The question Hilbert and Ackermann pose

[1] See [Gödel, 1986].

[2] All reprinted in [Gödel, 1986].

[3] See [Hilbert and Ackermann, 1972], though the question was anticipated as far back as 1837 by Bolzano in his *Wissenschaftslehre*, see [Bolzano, 1992].

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

is whether a certain axiom system for the first order predicate calculus '... is complete in the sense that from it all logical formulas that are correct for each domain of individuals can be derived.'[4]

Given that the relevant concepts had long been in place by then, and given that proving the theorem presents comparatively little in the way of mathematical difficulties, it is surprising that the question was stated as late as 1928. Even more surprising is the failure on the part of Hilbert and Ackermann and other contemporary logicians to notice that the theorem follows, in Gödel's words, 'almost trivially' from Skolem 1923 — which in turn builds on Löwenheim 1915 and Skolem 1920.[5] The reasons for this are interesting and accordingly the matter will be discussed below. Van Heijenoort and Dreben remark[6] that 'Throughout much of the 1920s it was not completeness but the decision problem for quantificational validity, a problem originating from the work of Schröder and Löwenheim, that was the dominant concern in studying quantification theory.' Perhaps this serves as a partial explanation. Examples of such results would include the semantic completeness of the propositional calculus due to Post in 1921 as well as a more general completeness theorem for the same due to Bernays in 1926, as well as the decision procedure for the first order monadic predicate calculus due to Behmann.[7]

Gödel's proof of the Completeness Theorem

We outline Gödel's proof as given in his 1929 thesis.[8] The deductive system is that of Hilbert and Ackermann 1928.

Gödel begins by giving a precise definition of all the terms involved. In particular the terms 'contradictory' and 'consistent' are defined in syntactic terms, by reference to derivability of a contradiction, rather than in semantic terms. This is an essential point of difference with Löwenheim 1915 and Skolem 1920 and 1923. As for the word 'provable,' 'the rules of inference yield (its) meaning' (a remark Gödel omitted from the subsequent article based on his thesis). A formula is valid if it is true in every model. The case in which identity is taken as part of the language is treated separately, as is the case in which the set of formulas whose satisfiability is considered is (countably) infinite.[9] The Completeness Theorem is as follows:

THEOREM 1. *Every valid formula of the restricted functional calculus is provable.*

But Gödel will actually prove the equivalent:

THEOREM 2. *Every formula of the restricted functional calculus is either refutable or satisfiable,*

[4]This translated quote is from [van Heijenoort, 1967], p.48.

[5]All English translations of the papers of Skolem and of Löwenheim cited here are from [van Heijenoort, 1967] See also [Skolem, 1970].

[6]See their introduction in [Gödel, 1986], p.47.

[7]See [Behmann, 1922].

[8]Reprinted and translated in [Gödel, 1986], to which we refer also for Gödel 1930.

[9]This latter extension was given in Skolem 1920 also, see below.

where a formula is 'refutable' if its negation is provable. Gödel defines the degree of a formula in prenex normal form beginning with universal quantifiers, to be the number of alternating blocks of quantifiers. He proves that if every formula of degree k is either satisfiable or refutable then so is every formula of degree $k+1$. This is not particularly difficult, and it is in this part of the proof that Gödel uses Skolem Normal Form.[10] It then remains to prove the assertion for formulas F of degree 1, i.e. of the form:

$$\forall x_1 \ldots \forall x_r \exists y_1 \ldots \exists y_s A(x_1, \ldots, x_r, y_1, \ldots, y_s) \tag{1}$$

For ease of exposition we sketch the proof in the special case of a sentence $(P)A = \forall x_1 \exists y_1 A(x_1, y_1)$, with $A(x_1, y_1)$ quantifier free, instead of the more general (1). Gödel defines a sequence of formulas

$$(P_n)A_n = \exists x_0 \exists x_1 \exists x_2 \ldots \exists x_n (A(x_0, x_1) \wedge (A(x_1, x_2) \ldots \wedge A(x_{n-1}, x_n)).$$

He then proves, by induction on n, that $(P)A \rightarrow (P_n)A_n$ is provable. This is a crucial lemma, as we explain below. Thinking of A_n as a propositional formula, with the atomic formulas considered as different propositional symbols, he then considers two cases:

Case 1: For some n, the propositional formula A_n is not satisfiable. In this case $\neg A_n$ can be derived in propositional logic, and hence in the (Hilbert-Ackermann) predicate calculus. Hence $\neg(P_n)A_n$ can be derived. But then $\neg(P)A$ is derivable, i.e. our original sentence is refutable, as $(P)A \rightarrow (P_n)A_n$ was derivable.

Case 2: For all n, the propositional formula A_n is satisfiable by some assignment s_n of truth-values to the atomic formulas. Let us note that, owing to the special form of A_n, the restriction $s_n \restriction m$ of s_n to the atomic formulas of A_m is a satisfying assignment for A_m. Now Gödel forms a tree of all possible assignments satisfying the various A_n. A node s satisfying A_n is defined to be an extension of each of its restrictions $s \restriction m$. An assignment of A_n is said to be on level n. This tree is finitely branching since there are only two truth-values and only finitely many atomic formulas in any single A_n. Since there is, by assumption, an assignment on each level of the tree, the tree is infinite. By an argument familiar nowadays from the textbook proof of König's Lemma[11] Gödel concludes that there must be an infinite branch $s_0 < s_1 < \ldots$ in the tree. Taken together the assignments or nodes of this branch form a single assignment satisfying each A_n. From this assignment one readily defines a model of the original sentence $(P)A$, and thereby $(P)A$ is satisfiable.

Building the tree T in the general case requires a carefully chosen order in which all sequences are handled. Regarding the general case also it should be pointed

[10] This is relevant to the discussion about Gödel and Skolem below.

[11] Later Gödel told Wang that at the time he did not know this [then only recently published] lemma [Wang, 1996, p.82]. Incidentally, in that conversation with Wang Gödel mistakenly identifies Dénes König, who formulated the Lemma, with his father Julius, who had worked on the continuum problem.

out that the crucial lemma asserting the provability of $(P)A \rightarrow (P_n)A_n$ is the step missing from the earlier versions of the proof due to Löwenheim and Skolem. Essentially the lemma builds the bridge from semantics back to provability.

We note that the Löwenheim-Skolem Theorem, which states that the model in the satisfiability case can always be taken to be countable, is a corollary of Gödel's Completeness Theorem.

In 1930 Gödel published the paper based on his thesis, notable as an extension of his 1929 mainly for the statement of the compactness theorem (only implicit in his 1929): a countably infinite set of quantificational formulas is consistent if and only if every finite subset of it is consistent.[12]

The compactness theorem, one of the two or three main tools in (the then fledgling subject of) model theory, seems not to have drawn much interest at the time. As it turns out, compactness is quite a strong property for a logic to have, in the sense that no proper extension of first order logic satisfies both the compactness and Löwenheim-Skolem theorems.[13]

The formal vs. the informal notion of decidability

In the introduction to his thesis, Gödel observes that one might advance the following objection to his Completeness Theorem: doesn't the use of the law of excluded middle in its proof 'invalidate the entire completeness proof'?[14] The Completeness Theorem asserts 'a kind of decidability,' namely every quantificational formula is either provable or a counterexample to it can be given, whereas 'the principle of the excluded middle seems to express nothing other than the decidability of every problem.' Thus the proof may be circular: one uses the decidability of every question in order to prove just that assertion. But, Gödel remarks, what he has shown is the provability of a valid formula from 'completely *specified, concretely enumerated* inference rules,'[15] not merely from all rules imaginable; whereas the law of excluded middle is used informally (in the metalanguage as one would say nowadays), meaning that the notion of decidability or solvability asserted by the law is left unspecified. As Gödel puts it:

> ... what is affirmed (by the law of excluded middle) is the solvability not at all through specified means but only through all means that are *in any way imaginable* ... [16]

The step forward provided by Gödel is therefore that if we assume solvability by all means imaginable, then we have, in the case of a sentence of first order predicate

[12]The theorem was extended to the case of uncountable vocabularies by Maltsev in 1936, from which the Upward Löwenheim-Skolem theorem immediately follows. See [Mal′cev, 1971].

[13]This is a result due to Per Lindström, see [Lindström, 1969].

[14]See p.63 of [Gödel, 1986].

[15]Italics Gödel's.

[16]Gödel remarks in a footnote to this passage that the notion of provability by any means imaginable is perhaps 'too sweeping.' Nevertheless, this does not affect the basic distinction that Gödel wishes to make, between the formal and informal notions of provability.

calculus, solvability by very specific means laid out carefully beforehand.[17]

Forerunners

Noteworthy in connection with the Completeness Theorem are the cases of Löwenheim and of Skolem. In a 1920 paper entitled 'Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen,' Skolem proved what is now called the Löwenheim-Skolem Theorem, a kind of semantic version of the Completeness Theorem:[18]

THEOREM 3. *Every first order formula of the restricted predicate calculus is either not satisfiable, or is satisfied in a countable domain.*

The statement of the Löwenheim-Skolem Theorem differs from that of Gödel's Completeness Theorem. With soundness, the latter states that provability in a first order theory is equivalent to truth in all of its models; whereas the former says nothing about provability. Still the argument Löwenheim and Skolem use does not differ in its mathematical essentials from the satisfiability part of the theorem in Gödel's dissertation. Skolem first proves that quantificational formulas in what is now known as Skolem normal form, i.e. of the form $\Pi\Sigma R$, where R is quantifier free and Π and Σ represent blocks of universal and existential quantifiers respectively, constitute a reduction class for first order predicate calculus. (This means that corresponding to every quantificational formula F there is a formula G in Skolem normal form, such that F is satisfiable in a given domain if and only if G is.)

He then constructs a tree whose finite levels consist of satisfying assignments ordered in the appropriate way. In the case that the formula is satisfiable, Dedekind's theory of chains is used to obtain an infinite branch through the tree and hence a countable domain in which the formula is satisfied. Otherwise at some finite level of the tree there is no satisfying assignment, and then the formula is not satisfiable in any domain, countable or uncountable.

Skolem 1920 is a 'clarification' or 'simplification' (in van Heijenoort's and Skolem's words respectively) of Löwenheim's 1915 'Über Möglichkeiten im Relativkalkül.'[19] Löwenheim's is apparently the first paper to state and prove the Löwenheim-Skolem Theorem. In Löwenheim normal form the blocks of existential quantifiers are infinite in length[20] — they quantify over the domain element by element, so to speak, considerably complicating the exposition.[21] Löwenheim

[17]For another interesting argument in the introduction to Gödel's thesis see below.

[18]The paper originally appeared in [Skolem, 1920].

[19]A 'pioneer work in logic,' according to van Heijenoort. See [van Heijenoort, 1967], which contains a translation. It appeared originally in [Löwenheim, 1915].

[20]Recently Badesa has argued that contrary to appearances, Löwenheim's proof does *not* make use of infinitary formulas. See [Badesa, 2004].

[21]The real defect introduced by using infinitary formulas though, according to Skolem, is the fact that Löwenheim thereby 'takes a detour through the infinite.' Skolem's objection is in keeping with his finitism, which we discuss below.

proves that the normal form formulas represent a reduction class, and then proves the Löwenheim-Skolem Theorem for first order logic. The proof proceeds by the same method as given in Skolem 1920; that is, by constructing a tree of satisfying assignments which is either finite, in which case the formula is not satisfiable, or infinite, in which case a satisfying assignment can be obtained from the tree. A drawback of the proof and the second difference with Skolem's clarification is that Löwenheim simply asserts the existence of an infinite branch through the tree in the satisfiability case. The type of reasoning now familiar from König's lemma is missing.

The second paper of Skolem's in which the Löwenheim-Skolem Theorem can be found dates from 1922, and is based on a lecture entitled 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre,' given to the Fifth Congress of Scandinavian Mathematicians in Helsinki in 1922.[22] Here Skolem reproves the Löwenheim-Skolem Theorem for the predicate calculus, using a construction which anticipates to a somewhat greater degree the satisfiability part of Gödel's Completeness Theorem. This time Skolem was able to eliminate the use of the axiom of choice in the theorem, which he saw as an improvement. The paper emphasizes what Skolem called the 'relativity in set theory,' and is accordingly entitled 'An Observation on the Axiomatic Foundations of Set Theory.' 'Relativity in set theory' is now known as the Skolem paradox: categoricity fails for Zermelo-Fraenkel set theory, i.e. it has countable (as well as uncountable) models.[23] Skolem remarks that he had pointed relativity out to Bernstein already in the winter of 1914-15 (this was presumably after seeing Löwenheim's paper, though the dates allow for the conversation to have taken place also before the publication of Löwenheim 1915), but did not publish his observation to Bernstein before 1923 as:

> ... first, I have in the meantime been occupied with other problems; second, I believed that it was so clear that axiomatization in terms of sets was not a satisfactory ultimate foundation of mathematics that mathematicians would, for the most part, not be very much concerned with it. But in recent times I have seen to my surprise that so many mathematicians think that these axioms of set theory provide the ideal foundation for mathematics; therefore it seemed to me that the time had come to publish a critique.[24]

Skolem's 1928 lecture 'Über die Mathematische Logik' comes closer to a proof of the Completeness Theorem in various ways.[25] A proof procedure is given for which it can be shown, if one includes his earlier results, that it is sound and complete.

[22] The conference proceedings were published the following year, see [Skolem, 1923]. Skolem's paper was reprinted in [Skolem, 1970] and a partial translation can be found in [van Heijenoort, 1967].

[23] A theory is said to be categorical if it has only one model up to isomorphism; it is λ-categorical if it has only one model of cardinality λ, up to isomorphism.

[24] [van Heijenoort, 1967], p. 300.

[25] The lecture was delivered to the Norwegian Mathematical Society on October 22, 1928 and originally published in [Skolem, 1928].

Specifically it gives an effective procedure for determining for a given formula F in prenex normal form, for each n whether or not there is a satisfying assignment at the nth level of the tree. If there fails to be a satisfying assignment at some level of the tree, then $\neg F$ is valid. Noteworthy is the fact that the proof procedure Skolem has in mind is in modern terminology cut-free and has the subformula property, anticipating Herbrand and Gentzen.

This brings us to Herbrand's 1930, which we will treat in the context of the finitary viewpoint.

Herbrand's Completeness Theorem in relation to Finitism

As mentioned the question whether first-order logic is complete had been introduced in 1928 in the *Grundzüge der theoretischen Logik* by Hilbert and Ackermann. They advocated the finitary standpoint, which at the time was the dominant view in the foundations of mathematics. This circumstance has led Hao Wang to make an interesting observation on the interaction between philosophical views and mathematical thought. Wang first points out that 'the very concept of completeness involves the nonfinitary component of being true in "arbitrary domains" '; this is non-finitary because there is no finitary construction method to construct all possible domains. He then continues:

> Viewed from such a perspective, what is surprising is not how late the concept was introduced, but that it was introduced at all. To use a fashionable term, one could say that the 'paradigm' [namely, the finitary standpoint] at that time could not accommodate either the formulation of the problem or its solution. While it is plausible to say that G[ödel] initiated a new paradigm, the fact that the problem was raised in [Hilbert and Ackermann [1972]] which belonged to the old paradigm, shows that paradigms are not as rigid or insular as many enthusiasts of 'incommensurable paradigms' have suggested.[26]

Herbrand's notion of finitary reasoning, which he called 'intuitionistic reasoning', a conflation characteristic of the times, is as follows:

> By an intuitionistic argument we understand an argument satisfying the following conditions: in it we never consider anything but a given finite number of objects and of functions; these functions are well-defined, their definition allowing the computation of their value in a univocal way; we never state that an object exists without giving the means of constructing it; we never consider the totality of all the objects x of an infinite collection; and when we say that an argument (or a theorem) is true for all these x, we mean that, for each x taken by itself, it is possible to repeat the general argument in question, which

[26] See p. 269 of [Wang, 1987].

> should be considered to be merely the prototype of these particular arguments.[27]

It is clear that neither Löwenheim's 1915 completeness proof nor that of Skolem 1920 falls under Herbrand's notion of a finitary proof. This is the first objection to their proofs stated by Herbrand in his 1930. The second objection has to do with their use of normal, or what he calls 'canonical,' forms,[28] and the third is the following:

> Löwenheim and Skolem implicitly assume that, once we have a model of $\neg P$, P cannot be provided; but this means that the consistency of mathematics (or at least of arithmetic) is implicitly assumed; our theorem, on the contrary, will allow us to investigate the consistency of arithmetic.[29]

A remark in another paper of Herbrand's sheds further light on this:

> We could say that Löwenheim's proof was sufficient in mathematics; but, in the present work, we had to make it 'metamathematical' ... so that it would be of some use to us.[30]

Herbrand's emphasis here on the distinction between a mathematical and a metamathematical view of Löwenheim's proof is reminiscent of a remark of Gödel's to Wang about Skolem, in a letter of March 7, 1968:

> That he [i.e. Skolem] used non-finitary reasoning for Loewenheim's Theorem proves nothing, because pure model theory, where the concept of proof does not come in, lies on the borderline between mathematics and metamathematics, and its applications to special systems with a finite number of axioms actually belongs to mathematics, at least for the most part.[31]

Gödel remarks above that in pure model theory the concept of proof does not 'come in,' meaning perhaps that the objects with which one is principally concerned in pure model theory are structures rather than proofs. Correspondingly, then, in proof theory[32] the notion of 'truth' or 'model' does not 'come in.' In Herbrand's particular case, he simply identifies the truth of a quantificational

[27] See pp.288–289, footnote 5, of [Herbrand, 1971].

[28] An objection hard to make sense of since the use of those forms do not impose a restriction on the generality of those proofs.

[29] P.289, [Herbrand, 1971].

[30] See p. 176 of [Herbrand, 1971].

[31] See [Gödel, 2003b], p.494. Incidentally, in the preceding line, Gödel had written that 'Moreover, what he [i.e. Skolem] tried to accomplish on p.29 of his paper of 1929 (see the last paragraph of my letter of December 7, 1967) evidently was to eliminate transfinite arguments from metamathematics (in a manner quite similar to Herbrand's).'

[32] Herbrand's thesis is called 'Investigations in Proof Theory.'

proposition with its provability in quantification theory. In a similar syntactic vein, a domain, or *champ*, is a finite 'collection of letters' and an 'infinite domain,' so to speak, is a finite domain together with some attached functions, propositional functions and logical values. 'Infinite domains' are a mechanism which yield a series of expanding finite domains. This is the machinery needed to carry out Herbrand's view of universal formulas: '... and when one says that an argument (or a theorem) is true for all these x, that means that for each x taken individually, it is possible to repeat the general argument in question, which should be considered nothing but the prototype of these particular arguments.' Thus Herbrand's is a finitary approximation to a full completeness theorem. In modern terms, it is as follows:

THEOREM 4. *A first order formula A is valid if and only if A has a Herbrand proof. More generally, if T is a universal theory, then* $T \models A$ *if and only if A has a Herbrand T-proof.*[33]

Herbrand's thesis presented ideas which were to become central in proof theory.[34] It includes a theorem very close to the cut-elimination theorem. The main theorem has a gap. This was repaired in an unpublished manuscript by Gödel,[35] in the sense that a weaker statement sufficient for the theorem was extracted and proved. The same proof of the weaker statement was discovered independently and published in a series of papers published in 1963 and 1964 by Dreben, Anderaa and Andrews.[36] The latter also gives counterexamples to Herbrand's original assertion.

We have seen that any full completeness proof for first-order logic in the original sense of the question has to be non-finitary. For Gödel, because of his platonism or objectivism, this transfinite character presented no problem at all.[37] But for Herbrand, the non-finitary notion of an arbitrary domain would not have been meaningful.[38]

For further discussion of the relation of Herbrand 1930, Skolem's three papers discussed above, as well as that of Löwenheim 1915 to Gödel 1929, see Hao Wang's introduction to Skolem's Selected Works, as well as Dreben and van Heijenoort's

[33]See 'On Herbrand's Theorem,' Sam Buss, math.ucsd.edu/ sbuss/ResearchWeb /herbrandtheorem/paper.pdf

[34]Though Kreisel was to remark of Herbrand's thesis in his review of [Dreben *et al.*, 1963]: 'There is no indication of a clear conception of the completeness problem in the thesis.'

[35]Van Heijenoort writes in his introduction to Herbrand's Écrits (translations ours):

> In 1939 Bernays wrote that the proof Herbrand gives of his Fundamental Theorem is 'difficult to follow' (Hilbert and Bernays 1939, p.158, note 1), without claiming that it is erroneous. In 1943 Gödel convinced himself that Herbrand's argument contains a gap; he wrote a note on this for private use, but didn't publish anything.

and adds a footnote (j) to the second sentence: 'This information was communicated to me by Gödel in the course of a conversation in Autumn 1963.' (p.8). The manuscript itself is in the Gödel Nachlaß, Arbeitshefte volumes IV and V.

[36]In particular, see [Dreben *et al.*, 1963].

[37]See his letter to Wang of December 7, 1967 [Gödel, 2003b, p.398].

[38]For a discussion of the philosophical and historical context of Herbrand's theorem see van Dalen's review of [Herbrand, 1971] in [van Dalen, 1974].

exhaustive discussion in [Gödel, 1986] already cited, as well as the various introductions to those papers by van Heijenoort et al., in 'From Frege to Gödel'.[39] See also Goldfarb's extensive introduction to Herbrand 1971.

Gödel's interaction with and account of his forerunners

Löwenheim and Skolem arrived at the Löwenheim-Skolem Theorem for the first-order predicate calculus in 1923, based on their respective 1915 and 1920 (with 1928 as a supplement). In the part of the Completeness Theorem where one constructs a model for a satisfiable sentence, their construction was more or less the same as appears in Gödel 1929. What then was Gödel's contribution to the Completeness Theorem? The Löwenheim-Skolem theorem was stated in semantic terms as we saw: a contradictory sentence is a sentence that is not satisfiable, rather than one leading to a contradiction by a finite derivation. No deductive system is given, nor is there any specification of the syntax. Thus the question of completeness (and therefore any proof of it) remains, at best, implicit. This is because the step from non-satisfiability of a formula to the provability of its negation cannot be given without specifying the rules of proof. Accordingly Dreben and van Heijenoort[40] see Gödel's contribution as tantamount to '... the replacement of an informal notion of 'provable' by a formal one ... and the explicit recognition that there is a question to be answered.' But this assessment misses an essential aspect of Gödel's contribution; it also fails to take Skolem's point of view sufficiently into account. For both Herbrand and Skolem the notion of 'all models' of a theory would not have been a meaningful one. I.e., there was not so much a lack of recognition at work as a conscious wish to avoid infinitary reasoning. Moreover, for Skolem his version of the completeness theorem (i.e. the Löwenheim-Skolem theorem) is first and foremost a theorem about the failure of categoricity in set theory, as was noted above. He drew from it what was to him the inevitable conclusion that Zermelo-Fraenkel set theory cannot be the right, *first order* way to formalize finitary reasoning.[41] This is because for Skolem, any theory which captured finitary reasoning in the right way would have to be first order, and would have to be categorical. Thus pursuing the syntactic aspect of the proof would have meant, it seems, taking a step in the opposite direction of what would have been the natural one for him.

As for Gödel's contribution to the Completeness Theorem, it was simply to solve the infinitary Hilbert-Ackermann question. This required seeing that both the refutation procedure and the satisfaction procedure can be incorporated into a single framework, as distinct from Löwenheim, Skolem and Herbrand, who, as we pointed out, saw these as cases to be handled separately, if at all. It should also be noted that Gödel solved the Hilbert-Ackermann question in such a definitive way

[39] [van Heijenoort, 1967]

[40] See [Gödel, 1986].

[41] For Skolem's account of this view see Skolem 1958 in [Skolem, 1970], a retrospective view of the 1920's.

as to dampen objections[42] concerning the infinitary aspects either of the statement of the theorem, or of its proof—objections which could easily have arisen in the general logical community, given the times.[43] And such was to become Gödel's trademark in his logical work: the vindication of infinitary methods in metamathematics by showing they can be handled in an exact way, that is to say: *mathematically*. Gödel put it this way:

> The Completeness Theorem, mathematically, is indeed an almost trivial consequence of Skolem 1923a. However, the fact is that, at that time, nobody (including Skolem himself) drew this conclusion (neither from Skolem 1922 nor, as I did, from similar considerations of his own ... This blindness (or prejudice, or whatever you may call it) of logicians is indeed surprising. But I think the explanation is not hard to find. It lies in the widespread lack, at that time, of the required epistemological attitude toward metamathematics and toward non-finitary reasoning.[44]

As an aside, completeness as such seems not to have been very much in the air in the early and mid-1920's. For example Fraenkel in his 1923 review of Skolem 1923 does not mention completeness at all, never mind that it might follow from Skolem's paper.[45] Also, the very fact that completeness is stated as an open question in Hilbert and Ackermann 1928 implies that, if Hilbert and Ackermann read the relevant papers of Löwenheim and Skolem before they published their text, they obviously did not obtain from them a solution of their problem.

As a final word on the interesting history of the Completeness Theorem (a very brief account of which has been given here), the Löwenheim-Skolem Theorem as given especially in Skolem 1923, has raised a number of questions as to when Gödel might have come across that paper. If one might speculate that it was not known to Hilbert and Ackermann, as we do above, then one might speculate in turn that it was also not known to Gödel. Library slips from the Gödel Nachlaß(which have been reexamined just recently) indicate that Gödel requested the proceedings volume containing Skolem 1923 once in November 1928, before he turned in his dissertation in July 1929, and twice afterwards. He was not able to obtain Skolem 1923 from any of the libraries he requested the volume from before November 1929 at the earliest, though. Indeed the proceedings volume which contains the paper was not owned by the National or University libraries in Vienna at the time (nor do they own it now).[46]

[42] At least the objections of those whose view of the issue was not already firmly fixed in the finitist camp.

[43] As we have seen, Löwenheim's version of the theorem had been seen as far from definitive.

[44] Letter to Wang, Dec. 7, 1967. See [Gödel, 2003b].

[45] Indeed Skolem himself, in his 1938 review of the second edition of Hilbert-Ackermann, notes their inclusion of theorems proved since the previous 1928 edition appeared; among these new theorems he includes the Completeness Theorem.

[46] See van Atten, 'On Gödel's awareness of Skolem's Helsinki lecture,' [van Atten, 2005].

One could also wonder about Herbrand's contact with Skolem 1923. Van Heijenoort writes,

> Herbrand himself writes (see below, p.217) that he has found his method 'by taking up again an idea of Loewenheim's'. As the many references in his writings indicate, Herbrand knew Loewenheim 1915 very well. Let us indicate what seems to have been Herbrand's readings in logic. The first book on logic he read seems to be Zaremba 1926. Next he read Principia Mathematica (first edition), at least the first volume, the papers of the German school (Hilbert, von Neumann, Bernays, Ackermann), Hilbert and Ackermann 1928. He read Gödel 1931 when it appeared. It is difficult to say whether he knew Skolem 1920 first hand or through the references to it given by Hilbert and Ackermann. He never knew, it seems to me, Skolem 1928 and 1929. There is in his writing no reference at all to Gödel 1930, which he cannot have failed to know.[47]

Van Heijenoort says nothing about Skolem 1923; presumably Herbrand did not know it.

Selected Observations and Corollaries

Categoricity fails also for other theories. For example, the existence of non-standard models of the Peano axioms follows from completeness (together with compactness).[48] This fact was not pointed out by Gödel in any of the publications connected with the Completeness Theorem from that time, and it seems not to have been noticed by the general logic community until much later. Skolem's definable ultrapower construction from 1933 gives a direct construction of a non-standard model of True Arithmetic (which of course extends Peano Arithmetic (henceforth PA)).[49] But Skolem never mentions that the existence of such models follows from the completeness and compactness theorems. Gödel in his review Skolem of 1933 also does not mention this fact, rather observing that the existence of non-standard models of True Arithmetic follows from the *incompleteness* theorem. One obtains the non-standard models more directly from incompleteness than from completeness and compactness, so perhaps this is natural.

On the flip side of the coin so to speak, completeness emphasizes the uniformity of the models of a given theory: if a first order sentence holds in all models, it must be because of a uniform reason (proof) and not for an accidental reason

[47] See p.3, footnote b of van Heijenoort's introduction to [Herbrand, 1968].

[48] These are models with infinitely large integers in them. One constructs these models as follows: add a new constant symbol c to the language of arithmetic. Extend P to a new theory P^* by adding to it the infinite collection of axioms: $\{c > \underline{0}, c > \underline{1}, \ldots\}$. P^* is finitely consistent hence consistent, hence by the Completeness Theorem it has a model.

[49] True Arithmetic is the set of arithmetic sentences true in the natural numbers.

which happens in each model or in each case for a different reason.[50]

Finally the Completeness Theorem gives the practitioner of informal rigor a powerful new tool: to check whether a sentence is provable in a theory, one needs only to check whether the sentence holds in an arbitrary model of the theory.

An Observation in the Introduction to Gödel's Thesis

Before going on to the Incompleteness Theorem, we take note of an observation Gödel makes in the introduction to his 1929 thesis. It would seem to be the very first occasion on which Gödel aims to prove an assertion which is metaphysical in nature from mathematical considerations. The argument was not included in the subsequent article Gödel published, based on the thesis. Apart from its content, its interest lies in showing that Gödel had gone some way toward thinking about incompleteness already by 1929.

As background, Gödel is reacting to a tenet of formalism, that consistency is a sufficient ground for existence. Another way of putting this is, one needs only to establish the *consistency* of an axiom with one's preferred axiom system, say, Zermelo-Fraenkel set theory, to infer the existence of those objects to which the axioms refer. Gödel argues against this assertion as follows:

> L.E. Brouwer, in particular, has emphatically stressed that from the consistency of an axiom system we cannot conclude without further ado that a model can be constructed. But one might perhaps think that the existence of the notions introduced through an axiom system is to be defined outright by the consistency of the axioms and that, therefore, a proof has to be rejected out of hand. This definition (if only we impose the self-evident requirement that the notion of existence thus introduced obeys the same operation rules as does the elementary one), however, manifestly presupposes the axiom that every mathematical problem is solvable. Or, more precisely, it presupposes that we cannot prove the unsolvability of any problem. For, if the unsolvability of some problem (in the domain of real numbers, say) were proved, then, from the definition above, there would follow the existence of two non-isomorphic realizations of the axiom system for the real numbers, while on the other hand we can prove the isomorphism of any two realizations ...

For emphasis we restate the argument: suppose Hilbert's principle — consistency implies existence—is true. Then assuming that there is a proposition ϕ about the reals, such that both ϕ and $\neg\phi$ are consistent with Hilbert's second order axiomatization, it then follows that there must be two non-isomorphic models M and N of those axioms, by the principle. But Hilbert's second order axiomatization of

[50] This observation is due to Andreas Blass, see 'Some Semantical Aspects of Linear Logic,' [Blass, 1998].

the reals is categorical,[51] meaning there is only one model up to isomorphism. We thus obtain a contradiction. So Hilbert's principle must be false. Note that the existence of M and N would follow also from Gödel's own Completeness Theorem, if a first order axiomatization of the reals were used. But then we would not have categoricity and no contradiction would follow.

It is at this point important to observe that we made above also the assumption of the existence of a proposition ϕ such that both ϕ and $\neg\phi$ are consistent with the axioms. Should we not abandon this assumption rather than Hilbert's Principle, that is, isn't this a proof of the completeness of Hilbert's axiomatization? The answer turns out to be no, although it is puzzling how Gödel could have known this. We know now that an axiomatization of the reals cannot be at the same time effective and categorical, as follows from Gödel's Incompleteness Theorem. Thus no effective second order axiomatization of the reals can be complete, unless we resort to the non-standard models of Henkin, in which case, however, we lose categoricity. Interestingly enough, Gödel could not have known this without knowing his own incompleteness theorem, as such an effective categorical axiomatisation would induce an effective complete axiomatization of number theory, ruled out by First Incompleteness Theorem. So either he was already close to the incompleteness theorem and the related observation in 1929, or he simply made a guess that the unsolvability of some problem about the reals could be proved. Such a guess would have been a reasonable one to make in 1929, as a considerable amount of discussion about the possible independence of, for example, the continuum hypothesis had taken place by then among for example Tarski, Bernays, Skolem and others.[52] The independence would of the continuum hypothesis would not be shown until 1963.

1.3 The Incompleteness Theorems

> Finally, we must also consider that it was not the controversy regarding the foundations of mathematics that caused the problem treated here[53] to surface (as was the case, for example, for the problem of the consistency of mathematics); rather, even if it had never been questioned that 'naive' mathematics is correct as to its content, this problem would have been meaningfully posed within this naive mathematics (unlike, for example, the problem of consistency) ... [54]

How did Gödel arrive at his undecidability results? As we saw, Gödel ruminated over the possibility that a question about the reals might be unsolvable already in the introduction of his thesis; a possibility which, he argued there, may constrain

[51] This fact was claimed by Hilbert about his axiomatization in 'Über den Zahlbegriff,' and there are indications of a proof in unpublished lecture notes from the late 1890's. By the late 1920's it had become 'folklore.'

[52] See [Kennedy and van Atten, 2004].

[53] The completeness of the first order predicate calculus.

[54] See [Gödel, 1986], p.65.

the principle that consistency is a criterion for existence. Showing the consistency of analysis was a key desideratum of the Hilbert program, and accordingly it was Gödel's turn to this question which led him to the two incompleteness theorems; two lectures of Brouwer, given in Vienna in 1928 also had a significant influence on Gödel's thinking of incompleteness.[55] We now briefly discuss the Hilbert program.

The Hilbert Program

The origins and aims of the Hilbert program are complex. Nevertheless we will describe here only those of its aspects which impinge directly on our discussion. The literature on the Hilbert program is extensive and we refer the reader to it.[56]

In the above quote Gödel refers to the controversy regarding the foundations of mathematics. This had arisen already in the nineteenth century in response to Cantor's theory of transfinite numbers, if not before, and was exacerbated by various set-theoretic and linguistic paradoxes which emerged later. Hilbert's proposal was that the axiomatic method should be used to effect finitary reductions of mathematical theorems and proofs, whenever these refer to infinite entities or use infinitary reasoning. From his 1925 'On the Infinite':

> There is, however, a completely satisfactory way of avoiding the paradoxes without betraying our science.
>
> 1. Wherever there is any hope of salvage, we will carefully investigate fruitful definitions and deductive methods. We will nurse them, strengthen them, and make them useful. No one shall drive us out of the paradise which Cantor has created for us.
>
> 2. We must establish throughout mathematics the same certitude for our deductions as exists in ordinary elementary number theory, which no one doubts and where contradictions and paradoxes arise only through our own carelessness.[57]

More precisely, Hilbert formulated the program[58] in (essentially) two parts. First, a set of indubitably self-evident axioms together with a finitary proof calculus are to be identified and expressed in a fixed language. This means that the symbols of the language are fixed and the axioms together with finitary rules of proof are stated explicitly in the language. The question of what constitutes a finitary rule of proof, or what constituted a finitary rule of proof then, e.g. for the Hilbert school, is not an uncontroversial one, as is the question of what is meant by an indubitably self-evident axiom. By a finitary system Hilbert could have meant,

[55] See the section on Gödel's results in intuitionistic logic and arithmetic below.

[56] Some of the main references are Sieg: [Sieg, 1990] and its sequel [Sieg, 1988], as well as [Sieg, 1999; Mancosu, 1998], and [Zach, Fall 2003].

[57] Originally appeared in [Hilbert, 1925] and was reprinted in e.g. [Benacerraf and Putnam, 1983].

[58] See for example [Hilbert, 1922], English translation in [Mancosu, 1998] and [Ewald, 1996]. For further references see [Zach, Fall 2003].

for example, what is now called Primitive Recursive Arithmetic.[59] The resulting system is a so-called formal system regarding which it is required that first, it is sound, i.e. any formal theorem provable in the system is true, and second, it is complete: meaning that any true mathematical statement should be provable in the system; and any false statement correspondingly should be refutable in the system.

The second part of the program asks for an elementary or finitary consistency proof for the system introduced in the first part.[60]

Carrying out the first part of the program would show that the 'ideal part' of mathematics, as they called it, namely the part which refers to infinite totalities, was conservative over what they called the 'contentual'[61] or 'real' part, i.e. the finitary part.[62] It would thereby reduce infinitary constructions and concepts to finitary ones.

Some formulations of the program were reductive in a strong sense: theorem-checking altogether should be reduced to scanning a finite number of slashes on both sides of an equality sign, checking whether their number was the same on both sides. This would mean that mathematical reasoning, in the end, rests merely on a capacity to distinguish one symbol from another. As Hilbert famously remarked: 'In the beginning was the sign.' Accordingly a slogan which came into use at the time attributed to the Hilbert school the aspiration to reduce mathematics to a 'formal game of symbols,' where again the reduction applies to semantic notions that might occur in a theorem such as truth or meaning, if not the mathematical objects themselves to which a theorem might refer, insofar as these are infinitary in nature.

It is important to note that the translation does not entail an *ontological* reduction, necessarily, and in fact the degree of commitment within the Hilbert school to an ontological reduction may not have been very pronounced.[63]

A contemporary expression of the thought which emphasizes the privileged role of finitism in mathematical reasoning is due to Tait, who sees that

> The special role of finitism consists in the circumstance that it is a minimal kind of reasoning presupposed by all nontrivial mathematical reasoning about numbers. And for this reason it is *indubitable* in a Cartesian sense that there is no preferred or even equally preferable ground on which to stand and criticize it. Thus finitism is fundamental to mathematics even if it is not a foundation in the sense that Hilbert wished.[64]

[59] See Tait's discussion of the thesis that the finitist functions are precisely the primitive recursive functions, a thesis he likens to Church's thesis, in 'Finitism,' reprinted in [Tait, 2005].

[60] The Hilbert program taken together with Gödel's theorems made urgent the question, whether elementary proofs are coextensive with finitary ones.

[61] 'Inhaltlich' in German.

[62] A theory A is a conservative extension of a theory B if any statement involving the lower system B which is provable in A can also be proved in B.

[63] See [Mancosu, 1998].

[64] [Tait, 2005], p. 22.

For Gödel, 'the Hilbert program, if successful, would have been of enormous epistemological value.'[65] The qualifier 'epistemological' is significant.

Expunging reference to infinite entities from mathematics will, if it succeeds, have the epistemological advantage of leaving a kind of mathematics that is evident to a higher degree.[66] But for Gödel there would be no concomitant ontological point to be drawn: mathematical objects are 'independent of our definitions and constructions.'[67]

The First Incompleteness Theorem provides a counterexample to completeness by exhibiting an arithmetic statement which is neither provable nor refutable in arithmetic. The Second Incompleteness Theorem shows that the consistency of (for example) arithmetic cannot be proved in arithmetic itself.

At first glance, Gödel's theorems would seem to demonstrate the unfeasibility of the Hilbert program. Von Neumann, for example, took this position.[68] In fact von Neumann went much further, asserting in a 1931 letter to Gödel:

> Thus, I think that your result has solved negatively the foundational question: there is no rigorous justification for classical mathematics.[69]

Gödel took a less drastic view of the situation both as regards classical mathematics and as regards the Hilbert program, though about the latter he would eventually come around to von Neumann's point of view.[70] As he wrote in his 1931,

> I wish to note expressly that theorem XI[71] ... does not contradict Hilbert's formalistic viewpoint. For this viewpoint presupposes only the existence of a consistency proof in which nothing but finitary means of proof is used, and it is conceivable that there exist finitary proofs that *cannot* be expressed in the formalism of P.

[65] Hao Wang in the 1960's

[66] See, for example, 7.1.9 on p.216 and 9.2.35 on p.302 of [**?**].

[67] See 'What is Cantor's Continuum Problem?', both the 1947 and 1964 versions as reprinted in [Gödel, 1990].

[68] As he wrote to Carnap in June of 1931 (Rudolf Carnap Nachlaß: RC 029-08-01, English translation in [Mancosu, 1999]):

> Thus today I am of the opinion that 1. Gödel has shown the unrealizability of Hilbert's program. 2. There is no more reason to reject intuitionism (if one disregards the aesthetic issue, which in practice will also for me be the decisive factor). Therefore I consider the state of the foundational discussion in Königsberg to be outdated, for Gödel's fundamental discoveries have brought the question to a completely different level. (Von Neumann is referring to the meeting on Logicism, Finitism and Intuitionism which took place in Königsberg in September of 1930, where Gödel announced his First Incompleteness Theorem, during a roundtable discussion on foundations.)

[69] [Gödel, 2003b], p. 339.

[70] See [Mancosu, 2004], p.117.

[71] Asserting the unprovability of consistency.

From the present perspective it would appear that the Hilbert program was not unfeasible. All that was needed is a wider notion of 'finitary.' We discuss relativised Hilbert programs below. Of course for some those relativizations represent too radical a diversion from the program in its original formulation. The reader is referred to the contemporary discussion of these matters. Also, Detlefsen[72] has argued, that because there are non-standard consistency predicates with respect to which the associated consistency statements *are* provable within the theory,[73] the second part of the program was not overthrown.

The First and Second Incompleteness Theorems

Under the influence of the Hilbert program then, in 1930 (if not already in 1929) Gödel turned to the problem of proving the consistency of analysis.

In his 'Logical Journey' Hao Wang published the full text of material Gödel had written (at Wang's request) about his discovery of the incompleteness theorems. This material had formed the basis of Wang's paper 'Some Facts about Kurt Gödel,' which was read and approved by Gödel:

> In the summer of 1930 I began to study the consistency problem of classical analysis. It is mysterious why Hilbert wanted to prove directly the consistency of analysis by finitary methods. I saw two distinguishable problems: to prove the consistency of number theory by finitary number theory and to prove the consistency of analysis by number theory ... Since the domain of finitary number theory was not well-defined, I began by tackling the second half... I represented real numbers by predicates in number theory... and found that I had to use the concept of truth (for number theory) to verify the axioms of analysis. By an enumeration of symbols, sentences and proofs within the given system, I quickly discovered that the concept of arithmetic truth cannot be defined in arithmetic. If it were possible to define truth in the system itself, we would have something like the liar paradox, showing the system to be inconsistent... Note that this argument can be formalized to show the existence of undecidable propositions without giving any individual instances. (If there were no undecidable propositions, all (and only) true propositions would be provable within the system. But then we would have a contradiction.)... In contrast to truth, provability in a given formal system is an explicit combinatorial property of certain sentences of the system, which is formally specifiable by suitable elementary means...

Rephrasing Gödel's remarks: He began by trying to reduce the consistency problem for analysis to that of arithmetic. But this required a truth definition

[72] See [Detlefsen, 1979].

[73] This is a result due to Feferman. See his 'Metamathematics in a General Setting,' [Feferman, 1960/1961].

for arithmetic, which in turn led to paradoxes, such as the Liar paradox: 'This sentence is false' and the Richard paradox, a paradox about definability. Gödel then noticed that if truth were replaced by provability, he could avoid the paradoxes. But this means that arithmetic truth, which is undefinable in arithmetic, and arithmetic provability, which is so definable, are not co-extensive. The First Incompleteness Theorem now follows, as the set of sentences provable in arithmetic is a (proper) subset of the set of arithmetic truths.[74] But, as Gödel wrote to Zermelo about this argument: 'This proof has, however, the disadvantage that it furnishes no construction of the undecidable statement and is not intuitionistically unobjectionable.'[75]

Judging from Gödel's account of the heuristics behind the theorem, it is apparent that he obtained a form of Tarski's 1935 theorem on the undefinability of truth in arithmetic in 1930.[76] He never published the result, though. The contemporary, vehement bias against semantics, biases which came to the fore, for example, when Tarski announced his results on the undefinability of truth in formal systems in 1935, served as a deterrent, perhaps, to Gödel's publicizing this result.

How does the definability of truth lead to paradoxes? We can think of any formula ϕ as defining a real as a set S of natural numbers, i.e. $x \in S \leftrightarrow \phi(x)$ holds in arithmetic. If one tries to interpret reals by definable sets of natural numbers then it is possible that one definition may code a lot of definitions of other reals - maybe even a definition of the real one is defining. In particular, if truth were definable, we would get a contradiction, as follows: define a sequence of reals S_m by stipulating that n is in the set S_m if and only if the mth formula $A(v)$ of arithmetic is true at $v = n$. Then consider the following formula $F(v)$: '$v \notin S_v$.' If this is the mth formula then m is in S_m if and only if m is not in S_m, a contradiction.

We now describe the proofs of the two theorems, formulating Gödel's results in PA, denoted in what follows by P. Gödel used the system defined in Principia Mathematica with the Peano axioms imposed on elements of type 0, the natural numbers. Gödel gives the informal argument first,[77] and then proves the first theorem (only) in complete detail. The informal argument is insufficient as it appeals to the notion of truth in an interpretation, a step contemporary logicians would have taken issue with, as we have noted. Accordingly such semantical notions as truth or meaning are expunged from the formal proof. As Gödel puts it, 'The purpose of carrying out the above proof with full precision in what follows is, among other things, to replace the second of the assumptions (every provable formula is true in the interpretation considered) by a purely formal and much

[74] For a contemporary account, see Gödel's correspondence with Zermelo, see pp.427 and 429 of [Gödel, 2003b]; see also Gödel's draft letter to Yossef Balas [Gödel, 2003a, pp.–11] and Wang's discussion in [Wang, 1987, pp.84–85].

[75] [Gödel, 2003b], p.429.

[76] See Gödel's 1931 correspondence with Zermelo and also with Bernays in p. 97 of [Gödel, 2003a], in which he outlines the truth definition in some detail.

[77] We take as our source for the two theorems Gödel 1931 as reprinted in [Gödel, 1986].

weaker one.'[78] Another, related reason for proving the theorem in detail is to insure that it is 'intuitionistically acceptable,"[79] as Gödel remarks in a 1931 letter to Zermelo, by which is meant 'finitistically unobjectionable.'[80] Yet another reason for carrying out the informal argument is to make more apparent its applicability to systems other than Principia.

But Gödel frequently relied on informal arguments. As we will see, for those logical theorems whose proofs are similar to the First Incompleteness Theorem,[81] Gödel never gave detailed proofs—if he gave any proof at all.

Before proceeding to the details of the formal proof, we define the notion of ω-consistency used by Gödel in the First Incompleteness Theorem. It is as follows: P is *ω-consistent* if $P \vdash \neg\phi(\underline{n})$ for all n implies $P \nvdash \exists x\phi(x)$. Naturally this implies consistency and follows from the assumption that the natural numbers satisfy the axioms of PA.

One of the main technical tools used in the proof is *arithmetization*, a mechanism which assigns natural numbers to terms and formulas of our formal theory P. This is nothing more than an injective mapping from the strings of symbols of our language into the natural numbers. The natural number corresponding to a formula ϕ is called its Gödel number and denoted by $\ulcorner\phi\urcorner$. Similarly we can assign a Gödel number to a sequence of formulas, most notably, a proof. The essential point here is that when a formula is construed as a natural number, then that natural number can occur as the argument of a formula, thus enabling the syntax to refer to itself. This will eventually allow Gödel to formalize the Liar paradox by substituting into the formula which says, 'the formula whose code is x is unprovable,' its own natural number code (or more precisely what is substituted is the formal term corresponding to the natural number).

Another concept required to carry out the formalization is the concept of numeralwise expressibility of number theoretic predicates. A number-theoretic formula $R(n_1, \ldots, n_k)$ is *numeralwise expressible* in P if for each tuple of natural numbers $(n_1, \ldots, n_k)$:

$$\begin{aligned} R(n_1, ..., n_k) &\Rightarrow P \vdash R(\underline{n}_1, ..., \underline{n}_k) \\ \bar{R}(n_1, ..., n_k) &\Rightarrow P \vdash \neg R(\underline{n}_1, ..., \underline{n}_k), \end{aligned}$$

where $\underline{n}$ is the formal term which denotes the natural number n.[82] One of the principal goals is to numeralwise express the predicate $\mathrm{Prf}(x, y)$: 'the sequence

[78] Insertion ours.

[79] See below for what Gödel means by 'constructively acceptable.' See also Tait's discussion of Gödel 1933b ([Gödel, 1986]) in 'Gödel's Unpublished Papers in Foundations of Mathematics,' [Tait, 2001].

[80] See also [Mancosu, 2004], where Gödel's 1931 letter to Zermelo is also used to illustrate this point.

[81] For example those which involve the construction of a fixed point, like the Speed-up Theorems.

[82] In P this is $S(S(\ldots S(0)\ldots))$, where n is the number of iterations of the successor function applied to the constant symbol 0.

with Gödel number x is a proof of the sentence with Gödel number y.' Reaching this goal involves defining forty-five relations we now call primitive recursive,[83] each defined in terms of the preceding ones. Relations needed are, among others, those which assert of a natural number that it codes a sequence, or a formula, or an axiom, or that it is the code, denoted by $Sb(r^{u_1 \cdots u_n}_{Z(x_1)\cdots Z(x_n)})$, of a formula obtained from a formula with code r by substituting for its free variable u_i the x_ith numeral for $i = 1, ..., n$. The forty-fifth primitive recursive relation defined is $\mathrm{Prf}(x, y)$, and the forty-sixth is $\mathrm{Prov}(x)$—'the sentence with Gödel number x is provable in P'—which without being primitive recursive, is however obtained from $\mathrm{Prf}(x, y)$ by existentially quantifying x. ($\mathrm{Prov}(x)$ satisfies only the 'positive' part of numeralwise expressibility, and not the negative part, but the negative part is not needed.)

In Theorem V of his paper, Gödel proves that any number theoretic predicate which is primitive recursive is numeralwise expressible in P. Thus since $\mathrm{Prf}(x, y)$ and substitution are primitive recursive, these are decided by P. This is the heart of the matter as we will see. Another key point about numeralwise expressibility is that although we informally interpret, for example, $\mathrm{Prov}(Sb(r^{u_1 \cdots u_n}_{Z(x_1)\cdots Z(x_n)}))$, by: 'the formula with Gödel number r is provable if the Gödel number for the x_ith numeral is substituted in place of the ith variable,' neither the formal statement within the theory P nor anything we prove about it appeals to such meanings. On the contrary $\mathrm{Prov}(Sb(r^{u_1 \cdots u_n}_{Z(x_1)\cdots Z(x_n)}))$ is a meaningless string of logical and arithmetical symbols. As Gödel puts it in his introduction to his theorem V, 'the fact that can be formulated vaguely by saying that every recursive relation is definable in the system P (if the usual meaning is given to the formulas of this system) is expressed in precise language, *without* reference to any interpretation of the formulas of P'.[84]

The First Incompleteness Theorem (Theorem VI in his paper) may be stated

THEOREM 5 Gödel's First Incompleteness Theorem. *If P is ω-consistent, then there is a sentence which is neither provable nor refutable*[85] *from P.*

By judicious coding of syntax referred to above, Gödel is able to write a formula $\mathrm{Prf}(x, y)$[86] of number theory, representable in P, so that

$$n \text{ codes a proof of } \phi \quad \Rightarrow \quad P \vdash \mathrm{Prf}(\underline{n}, \underline{\ulcorner \phi \urcorner}) \tag{2}$$

$$n \text{ does not code a proof of } \phi \quad \Rightarrow \quad P \vdash \neg \mathrm{Prf}(\underline{n}, \underline{\ulcorner \phi \urcorner}). \tag{3}$$

Let $\mathrm{Prov}(y)$ denote the formula $\exists x \mathrm{Prf}(x, y)$.[87] By an ingenious use of substitu-

[83] Gödel uses the word 'recursive' in place of primitive recursive. For a definition of the terms 'recursive' and 'primitive recursive' see [Rogers, 1967]. The use of recursive functions apparently goes back to Grassmann, see p. 147 of [Wang, 1957].

[84] [Gödel, 1986], p.171, italics Gödel's.

[85] A sentence is refutable from a theory if its negation is provable.

[86] Gödel uses the notation $B(x, y)$ for $\mathrm{Prf}(x, y)$.

[87] Gödel uses the notation $Bew(y)$ for $\mathrm{Prov}(y)$.

tion Gödel is able to write down a sentence ϕ with the property

$$P \vdash (\phi \leftrightarrow \neg \mathrm{Prov}(\underline{\ulcorner \phi \urcorner})). \tag{4}$$

This method is an instance of what is called Gödel's Fixed Point Theorem.[88] Thus ϕ says 'I am not provable.' We observe, if $P \vdash \phi$, then by (2) there is n such that $P \vdash \mathrm{Prf}(\underline{n}, \underline{\ulcorner \phi \urcorner})$, hence $P \vdash \mathrm{Prov}(\underline{\ulcorner \phi \urcorner})$, hence $P \vdash \neg\phi$, so P is inconsistent. Thus

$$P \nvdash \phi. \tag{5}$$

Furthermore, by (5) and (3), we have $P \vdash \neg \mathrm{Prf}(\underline{n}, \underline{\ulcorner \phi \urcorner})$ for all natural numbers n. By ω-consistency $P \nvdash \exists x \mathrm{Prf}(x, \underline{\ulcorner \phi \urcorner})$. Thus (4) gives $P \nvdash \neg\phi$. We have shown that if P is ω-consistent, then ϕ is independent of P.

On concluding the proof of the first theorem, Gödel remarks, 'we can readily see that the proof just given is constructive; that is... proved in an intuitionistically unobjectionable manner... '[89] This is because, as he remarks, all the existential statements are based on his theorem V (giving the numeralwise expressibility of primitive recursive relations), which is intuitionistically unobjectionable.

Before turning to the Second Incompleteness Theorem, Gödel proves Theorem X of the paper, which states that the undecidable sentence can be written by means of plus and times only. Later Gödel improved the result by showing that the undecidable sentence can even be taken to be of the form

$$\forall x_1, \ldots, x_n \exists y_1, \ldots, y_m R(x_1, \ldots, x_n, y_1, \ldots, y_m),$$

where R is Diophantine. He also shows that the class of such formulas, which he calls class A, is undecidable, presciently making use of the Pell equation and other machinery used later in the proof of the undecidability of Hilbert's Tenth Problem.[90] This result, known as the MRDP theorem,[91] states the undecidability of the simpler class of formulas of the form: $\exists y_1, \ldots, y_m R(y_1, \ldots, y_m)$, where R is Diophantine.

Let $\mathrm{Con}(P)$ be the sentence $\neg \mathrm{Prov}(\underline{\ulcorner 0 = 1 \urcorner})$.

THEOREM 6 Gödel's Second Incompleteness Theorem. *If P is consistent, then* $\mathrm{Con}(P)$ *is not provable from P.*

Let ϕ be as in (4). By the proof of (5) above, if $P \vdash \phi$, then $P \vdash 0 \neq 1$. The reasoning used to infer 'if $P \vdash \phi$, then $P \vdash 0 \neq 1$' does not go beyond elementary number theory, and can therefore be formalized in P. This yields: $P \vdash (\mathrm{Prov}(\underline{\ulcorner \phi \urcorner}) \rightarrow \neg \mathrm{Con}(P))$, and thus by (4), $P \vdash (\mathrm{Con}(P) \rightarrow \phi)$. Since $P \nvdash \phi$,

[88] Note that although Gödel constructs a fixed point in the course of proving the incompleteness theorem, he does not state the fixed point theorem explicitly.

[89] [Gödel, 1986], p.177.

[90] These results were found in an unpublished manuscript in Gödel's Nachlaß. Denoted 193? they were subsequently published in [Gödel, 1995]. The manuscript likely formed the basis for a lecture that was to be given at the 1940 ICM, an event which was cancelled due to the war.

[91] The theorem is due to Matiyasevic, Robinson, Davis and Putnam.

we must have $P \nvdash \mathrm{Con}(P)$. The bulk of the proof, which we omit here, is the step involving the formalization.

This concludes the proof of Gödel's Second Incompleteness Theorem. It is noteworthy that ω-consistency is not needed. Also note that neither is $\neg\mathrm{Con}(P)$ provable, by the consistency of P and the fact, now known as Löb's theorem, that $P \vdash \mathrm{Prov}(\ulcorner\phi\urcorner)$ implies $P \vdash \phi$.

The assumption of ω-consistency in the First Incompleteness Theorem was eliminated by Rosser in 1936, and replaced by the weaker notion of consistency. Rosser's generalization involves applying the fixed point theorem to the formula $R(x)$: 'for all z: either z is not the Gödel number of a proof of the formula with Gödel number x or there is a proof shorter than z of the negation of (the formula with Gödel number) x.' [92]

With regard to the Second Incompleteness Theorem, the argument relies in part on formalizing the proof of the First Incompleteness Theorem as we saw. This step is omitted in Gödel 1931. He planned to include the step in what would have been part II,[93] but instead of writing it he turned to the continuum problem.[94] (Part II of his 1931 was to elaborate on other points of his 1931 too: the 'true reason for incompleteness,' see below, and the applicability of the two theorems to other systems.) Having specified both the arithmetization and the numeralwise expressibility of the required primitive recursive functions in complete detail, he perhaps did not feel compelled to attend to what looked like an exercise in formalization, relying instead on the informal argument to convince (in which it succeeded). However this step turned out to be somewhat non-trivial.[95] Eventually a complete proof of the Second Theorem was given by Hilbert and Bernays in some seventy pages in their 1939. A much more succinct treatment of the theorem was given by Löb in his 1955, and subsequently Feferman, in his 1960 'Arithmetization of Metamathematics in a General Setting,'[96] gave a succinct and completely general treatment of both the First and Second[97] Theorems. Feferman's paper is also

[92] See [Rosser, 1936].

[93] See footnote 48a of Gödel 1931.

[94] 'Jetzt, Mengenlehre!'—and now, [on to] set theory!—Gödel is alleged to have said around that time (see p.128 of Hao Wang, Popular Lectures on Mathematical Logic, [Wang, 1981], as well as J. Dawson, Logical Dilemmas: The Life and Work of Kurt Gödel, [Dawson, 1997].) Gödel's interest in set theory may have begun to develop as early as 1928 when he requested at the library the volume containing Skolem's Helsinki lecture (as was mentioned earlier). It is very likely Gödel did not actually get the volume, see above and [van Atten, 2005]. Be that as it may, Dawson goes on to mention that in 1930 Gödel requested works by Fraenkel (Einleitung in die Mengenlehre, in which Gödel will have noticed Fraenkel's skepticism about Hilbert's attempted proof of the continuum hypothesis), von Neumann, and the paper in which Hilbert had put the problem of deciding the continuum hypothesis on the agenda of twentieth century mathematics.

[95] As Kleene puts it in his excellent 1986 introduction to Gödel 1931, of the informal presentation, 'Certainly the idea of the argument for Theorem XI (consistency) was very convincing; but it turned out that the execution of the details required somewhat more work and care than had been anticipated.' See pp.126-141 of [Gödel, 1986].

[96] [Feferman, 1960/1961]

[97] There is an interesting human story behind this theorem: Gödel had announced the First Incompleteness Theorem on September 7th, 1930, during a roundtable discussion that took place

noteworthy for his identification of non-standard consistency predicates with respect to which the associated consistency statement *is* provable, as was mentioned above.

---

on the penultimate day of a meeting in Königsberg, where he presented his thesis result on the completeness of first order quantificational logic. (The meeting is also notable for occasioning the famous retirement address of David Hilbert, 'Naturerkennen und Logik,' which ends with the impassioned assertion:

> Der wahre Grund warum es Comte nicht gelang, ein unlösbares Problem zu finden, besteht meiner Meinung noch darin, dass ein unlösbares Problem überhaput nich gibt. Wir müssen wissen; wir werden wissen. [Hilbert, 1965]
>
> (The real reason why Comte was not able to find an unsolvable problem, in my view, is that there are no unsolvable problems. We must know; we will know.)

Von Neumann attended the roundtable discussion, and sought Gödel out afterwards. A conversation ensued in which von Neumann pressed Gödel for the mathematical details of the latter's announced result. (In particular he asked if the undecidable sentence could be arithmetic, which Gödel would show in Gödel 1931.) Gödel returned to Vienna after the meeting and wrote the manuscript which was eventually to become Gödel 1931, submitting an abstract for it to the Vienna Academy of Sciences on October 23rd, 1930 through Hans Hahn, and submitting the full paper for publication on November 17th, 1930. A few days later, Gödel received a letter from von Neumann dated November 20th in which von Neumann wrote (see pages 336 and 337 of [Gödel, 2003b]):

> I have recently concerned myself again with logic, using the methods you have employed so successfully in order to exhibit undecidable properties. In doing so I achieved a result which seems to me remarkable. Namely, I was able to show the consistency of mathematics is unprovable.

Von Neumann went on to describe the informal proof of the result, 'by a somewhat different method' from that of Gödel's (as he wrote in his second letter to Gödel). Gödel, in a letter that has been lost, replied he had proved the second theorem already, and enclosed a reprint of Gödel 1930b (see [Gödel, 1986]). As 1930b does not contain a proof of the second theorem, Gödel must have given the argument for the theorem in his lost reply, as on November 29th von Neumann wrote to Gödel: 'As you have established the theorem on the unprovability of consistency as a natural continuation and deepening of your earlier results, I clearly won't publish on the subject.' It is a matter of occasional speculation whether the theorem would have been thought of as a joint one under the conventions governing such issues in the present day. Opinions vary somewhat on this point. The phrasing in von Neumann's reply—*clearly*—indicates that he may not have considered the step from the first theorem to the second to be so significant, as later commentators have held it to be (on the grounds that the formalizing the informal argument is non-trivial), as to warrant joint authorship. Relevant to the contemporary speculation perhaps is the fact that in both Gödel's and von Neumann's case, the proof that was given was an informal one, indeed more in the way of a heuristic than a proof. Perhaps conventions regarding rigor in mathematical logic proofs have shifted toward a more strict standard since that time.

Relevant to the overall question of when Gödel proved the Second Incompleteness Theorem is the fact that Gödel had discussed what appears to be *both* theorems with Carnap in the Cafe Reichsrat on the August 26th and 29th prior to the Königsberg meeting, see p. 86 of [Dawson, 1991]. One cannot be completely sure about the second theorem having been discussed as Carnap's diary notations concern Gödel's remarks on 'the difficulty of proving consistency'; Carnap seems not to have understood Gödel's remarks that day in any case, see p. 86 of [Dawson, 1991].

Two corollaries

Just as Gödel did not point out certain consequences of his completeness and compactness theorems, here too Gödel refrains from pointing out some corollaries of the First Incompleteness Theorem and the fixed point construction. The first we discuss is the undecidability of P. A theory is said to be *undecidable* if the set of Gödel numbers of its theorems is not recursive.

THEOREM 7. *P is undecidable.*

Proof. Suppose the set of Gödel numbers of theorems of P were recursive. Then by the representability of recursive relations[98] in P, there is a formula $\psi(x)$ of number theory such that

$$P \vdash \phi \Rightarrow P \vdash \psi(\ulcorner \phi \urcorner), \tag{6}$$

$$P \nvdash \phi \Rightarrow P \vdash \neg\psi(\ulcorner \phi \urcorner). \tag{7}$$

By a fixed-point construction applied to ψ there is a sentence θ with the property:

$$P \vdash (\theta \leftrightarrow \neg\psi(\ulcorner \theta \urcorner)).$$

Now observe that if $P \vdash \theta$, then (6) gives $P \vdash \psi(\ulcorner \theta \urcorner)$, hence $P \vdash \neg\theta$, so P is inconsistent. Thus $P \nvdash \theta$. But in this case by (7) $P \vdash \neg\psi(\ulcorner \theta \urcorner)$. Thus $P \vdash \theta$, contrary to the above. We have reached a contradiction and therefore we conclude that P is undecidable.[99]

■

A consequence of this theorem is a result due to Church, published in 1936, and is as follows:

THEOREM 8. *Predicate logic is undecidable.*

Proof. As is easily shown, if a number theoretic relation is numeralwise expressible in P, it is numeralwise expressible in a finitely axiomatized fragment of arithmetic such as Robinson's arithmetic Q. Thus we can replace P above by Q, and hence Q is also undecidable. Now suppose θ is the conjunction of the axioms of Q. For any sentence ϕ of number theory

$$Q \vdash \phi \iff \vdash (\theta \rightarrow \phi).$$

Thus predicate logic is decidable only if Q is; but Q is not decidable so therefore neither is predicate logic. ■

[98] Gödel used only the representability of primitive recursive relations in his 1931 but the representability of the wider class is easily shown.

[99] In fact, P is *essentially undecidable* in the sense that every consistent extension of P is undecidable.

This simple observation was published by Church in 1936.[100] The finite undecidable theory he constructed was different from Q.

A second corollary of the incompleteness theorems has to do with the existence of non-standard models of PA (denoted P in the above, as mentioned). Their existence follows already from the completeness and compactness theorems as we saw. But incompleteness (together with completeness as before) gives new kinds of non-standard models, those of the theory $PA + \mathrm{Con}(PA)$, or $PA + \neg\mathrm{Con}(PA)$. (Though of course the theory $PA + \mathrm{Con}(PA)$ also has among its models the standard model.) In fact this process of adding independent statements can be iterated into the transfinite. Taking all finite iterations yields the result that there are at least continuum many non-isomorphic countable non-standard models of arithmetic, as iterating the addition of both $\mathrm{Con}(T)$ and $\neg\mathrm{Con}(T)$ can be used to construct a full binary tree of consistent but inequivalent theories extending P. Although Gödel commented about systems which prove false theorems at Königsberg, he did not pursue the matter. The nature of these models is mysterious: a model of $P + \neg\mathrm{Con}(P)$ for example, contains a non-standard element[101] which is the Gödel number of a(n infinitely long) proof of $0 = 1$. As was proved later, if the addition and multiplication in a countable non-standard model of arithmetic are construed as ternary operations over the integers, such a model cannot be recursive.[102]

The 'true reason for incompleteness'

In footnote 48a of his 1931 paper Gödel expresses the real reason for the incompleteness of formal systems:

> As will be shown in Part II of this paper, the true reason for the incompleteness inherent in all formal systems of mathematics is that the formation of ever higher types can be continued into the transfinite (see Hilbert 1926, page 184), while in any formal system at most denumerably many of them are available. For it can be shown that the undecidable propositions constructed here become decidable whenever appropriate higher types are added (for example, the type ω to the system P). An analogous situation prevails for the axiom system of set theory.

Gödel's types are those in the sense of the type hierarchy given in Principia. He seems to be referring to the fact that the satisfaction predicate for a formal system can be defined in a formal system of higher type, which satisfaction predicate can then be used to decide Gödel sentences of the system of lower type.[103] What is

[100] 'A Note on the Entscheidungproblem,' reprinted in [Davis, 2004].

[101] I.e. an infinitely large integer.

[102] This is known as Tennenbaum's theorem, see for example [Kaye, 1991].

[103] For a detailed proof of this result see Craig, 'Satisfaction for n-th order languages defined in n-th order languages,' [Craig, 1965].

interesting about Gödel's observation is that it reduces the incompleteness of a theory to the barest of syntactical considerations: one needs only to look at what types the syntax allows, and then go up one level in the type hierarchy to decide the Gödel sentences.

Relativised Hilbert Programs

One could ask, given that PA does not admit of a finitary consistency proof, how should finitary reasoning be supplemented in order to prove the consistency of arithmetic? This was settled by Gerhard Gentzen in his landmark 1936 paper 'Die Widerspruchsfreiheit der reinen Zahlentheorie,'[104] in which he showed that what was needed was transfinite induction up to the countable ordinal ϵ_0, the countable ordinal obtained by taking the countable limit of the iterated exponential series ω, $\omega^\omega, \omega^{\omega^\omega}, \ldots$ The result is the best possible in that no smaller ordinal suffices; by coding one can speak about ordinals within arithmetic. One then proves that transfinite induction up to ordinals smaller than ϵ_0 is provable. The result inaugurated the era of ordinal analysis in proof theory, which, among other things, seeks to assign to a theory the least ordinal up to which transfinite induction is provable in the theory, and such that induction up to the ordinal proves the consistency statement for the theory. For an interesting discussion of Gödel's view that ϵ_0 is the precise limit of finitary reasoning see Tait's 'Finitism.'[105]

A kind of converse to Gentzen's theorem is known as the Turing Completeness Theorem: [106] Consider the sequence of theories P, $P + \mathrm{Con}(P)$, $P + \mathrm{Con}(P) + \mathrm{Con}(P + \mathrm{Con}(P)), \ldots$, iterated into the transfinite. Turing proved that at level $\omega+1$ the theory one obtains is Π_1-complete for arithmetic, meaning that the theory proves, not only the Π_1-sentence $\mathrm{Con}(P)$ added to it, but in fact all Π_1-sentences in the language of arithmetic which are true in the standard model. The theorem went some way toward 'bridging the gap,' as Feferman puts it, between recursive theories like PA and its highly non-recursive extension True Arithmetic[107]. This idea was considerably elaborated by Feferman in his 'Transfinite Recursive Progressions of Axiomatic Theories'[108] and opened up yet another branch of proof theory.

Another program inaugurated by Feferman and also by Kreisel in the spirit of a relativized Hilbert's program involves finding proof-theoretic reductions. This means showing for a given theory T the existence of a seemingly weaker theory T' over which the first theory is conservative. Feferman has argued that if T' is taken to be Primitive Recursive Arithmetic supplemented with Weak König's Lemma, then this meets the conservativity criterion of the Hilbert program, at least for all practical purposes; meaning that if one expands the notion of 'finitary' to include Weak König's Lemma, then the bulk of mathematics can be carried out within

[104] See [Gentzen, 1967]. English translation in [Gentzen, 1969].
[105] [Tait, 2005].
[106] See A.M. Turing, 'Systems of Logic based on Ordinals,' [Turing, 1939].
[107] True Arithmetic is the set of all sentences in the language of arithmetic true in the standard model.
[108] [Feferman, 1962, pp.259–316].

this theory. It is still incomplete but this does not affect conservativity.

Another outgrowth of the Hilbert program is the 'reverse mathematics' program of Simpson[109] and Friedman, which carries out the Feferman program piecewise, meaning the proof-theoretic reductions are given theorem by theorem.[110]

Finally, Gödel's Dialectica translation of 1958 gives a consistency proof for arithmetic. But this will be discussed at length below.

Other Developments: A Partial List

We mention just a few more theorems related to incompleteness. A comprehensive account of the theorems related to Gödel's incompleteness theorems will not be given here.

A natural question to ask after the publication of Gödel's 1931 results was, do self-reference phenomena exhaust incompleteness? That is, do all statements which are independent of the canonical systems we wish to consider involve some form of self-reference, or are there statements tied more closely to mathematical practice itself, which are also independent? The answer depends on the theory under consideration. While for set theory there is now a wide variety of seemingly natural statements which are independent, for arithmetic they are comparatively few. The first to appear was the landmark result from 1977 due to Jeff Paris and Leo Harrington,[111] who found a finite combinatorial statement independent of PA. Paris and Harrington called their paper 'A mathematical incompleteness in Peano Arithmetic' in order to emphasize the naturality of the statement, in contrast to the independent statements Gödel had found. Theirs is a generalized version of the finite Ramsey theorem, and a consequence of the Infinite Ramsey Theorem. It is independent because it is equivalent in PA to the uniform Σ_1-reflection principle for that theory.

It turned out to be difficult to obtain other natural independence results for arithmetic. Mostowski had already conjectured that this would be the case, because of their non-recursiveness, in his 'Thirty years of foundational studies':[112]

> All these results[113] explain why it is so difficult to establish independence of number-theoretical sentences from the arithmetical axioms by the use of nonstandard models: In order to find such applications one should have a much better knowledge of the structure of these models and more direct methods of their constructions than we have at present.

As for independence results in set theory there are many. The forcing method

[109] See, for example [Simpson, 1985].

[110] This subsection on Hilbert's program has relied in part on the Stanford Encyclopedia of Philosophy article 'Hilbert's Program' by Richard Zach, see [Zach, Fall 2003].

[111] 'A Mathematical Incompleteness in Peano Arithmetic,' see [Barwise, 1977].

[112] Lectures Mostowski gave in 1964, see [Mostowski, 1966].

[113] Referring to Tennenbaum type theorems, see above.

is nothing else but a method for producing such theorems. We discuss this further below, in the context of the question whether *mathematics* is incomplete.

The last development we mention concerns the hierarchy of weak arithmetic theories. For example, Robinson's finite arithmetic theory Q does not prove its own consistency statement; but surprisingly even the seemingly stronger theory $I\Delta_0$, which restricts the induction in Peano's axioms to formulas containing only bounded quantifiers, does not prove $\mathrm{Con}(Q)$.[114]

Is mathematics incomplete?

The incompleteness theorems make exigent the following question: are there meaningful or natural mathematical assertions which are absolutely undecidable, that is to say, undecidable by any proof the human mind could possibly conceive?[115] If one's notion of provability is restricted to provability in a particular formal system, then the answer is obviously yes—if one takes the Gödel sentences as meaningful. As it turns out, for Gödel provability is not to be restricted in this way. In a 1962 letter to Leon Rappaport Gödel wrote:

> I have not proved that there are mathematical questions undecidable for the human mind, but only that there is no *machine* (or *blind formalism*) that can decide all number theoretic questions (even of a certain very special kind).

As we have seen, already in his 1929 thesis Gödel separated the notion of the absolute provability of an assertion from the concept of its formal provability. In a number of papers written in the 1930's he continued this train of thought, with remarks on absolute undecidability vs. formal undecidability, absolute vs. formal consistency, and so on.[116] If the realist's minimal position holds that any meaningful mathematical proposition must admit of a yes or no answer, i.e., simply put, will (in principle) be decidable, then interestingly for the author of the First Incompleteness Theorem, Gödel settled on the minimal position by 1947, if not earlier.[117] Thus for Gödel, the incompleteness of certain formal systems, in particular the type of incompleteness which emerges from his 1931 method for producing independent statements of those formal systems, is merely, in the end, a residuum of formalization;[118] it does not emerge naturally in mathematical practice—at least insofar as that practice avoids the study of formal systems.

[114] See 'On the scheme of induction for bounded arithmetic formulas,' [Wilkie and Paris, 1987].

[115] This is Gödel's definition of absolute undecidability given in the Gibbs lecture, see [Gödel, 1995].

[116] For a detailed analysis of Gödel's position on absolute undecidability as it evolved from this period, see [Kennedy and van Atten, 2004].

[117] In the Grandjean questionaire, Gödel dates his Platonism from 1925: [Gödel, 2003a, pp.444, 447, 450].

[118] In set theory circles this is referred to as 'residual incompleteness.' See Dehornoy's 'Recent progress on the Continuum Hypothesis, after Woodin,' [Dehornoy, 2004].

A distinction that needs clarifying at this stage is that between statements that do play a role in mathematical practice and those that do not. The undecidable sentences constructed in the proofs of Gödel's incompleteness theorems are manifestly different from anything found in mathematical practice so far, and in that specific sense not mathematically meaningful. To a realist, the mathematical meaningfulness of a statement simply means that it has mathematical content (or is equivalent to one that does), in the sense that the terms in the statement refer.[119] To a (Hilbertian) formalist a certain statement may well be relevant to mathematics without being contentual (think of any practically relevant part of classical mathematics that is not finitary). On the other hand, this kind of background commitment associated with the realist and the formalist is often lacking in the colloquial use of the phrase 'mathematically meaningful' among mathematicians. That use rather emphasizes typical aspects (often of an aesthetic nature) such as being 'natural', 'fundamental', 'elementary', or 'interesting'.

For Gödel the continuum problem[120] is indubitably meaningful. In a 1947 paper entitled 'What is Cantor's Continuum Problem' [121] he describes it as a question 'from the "multiplication table" of cardinal numbers' and one of 'the most fundamental questions in the field.'[122]

> So the analysis of the phrase 'how many' unambiguously leads to a definite meaning for the question stated in the second line of this paper: The problem is to find out which one of the $\aleph$'s is the number of points of a straight line or (which is the same) of any other continuum (of any number of dimensions) in a Euclidean space.[123]

Gödel predicts in this paper, correctly as it turned out,[124] that the continuum hypothesis will be shown to be independent of the axioms of Zermelo-Fraenkel set theory. He argues forcefully in the paper that the continuum problem will nevertheless be solved, as it is a meaningful question. The independence, rather than precluding a solution of the problem, only shows that those axioms, though sufficient to decide almost any question from the core areas of mathematics like analysis and algebra, are insufficient to decide *all* meaningful mathematical questions. Thus one must find a proper extension of them, or possibly a series of extensions iterated into the transfinite, so as to decide in principle any meaningful mathematical question.

What sort of extension should one seek? In the 1946 'Remarks before the Princeton bicentennial conference on problems in mathematics'[125] Gödel mentions

[119] Gödel simply defines meaningfulness as being part of mathematics proper (or translatable into mathematics proper). See *1940a, [Gödel, 1995], p.176.

[120] I.e., the question as to the cardinality of the continuum

[121] Reprinted in [Gödel, 1990].

[122] See pp.256 and 257 respectively of [Gödel, 1990].

[123] [Gödel, 1990], p. 257.

[124] Though independence was not shown until 1961 with the work of Paul Cohen, see [Cohen, 1963].

[125] See pp.150-153 of [Gödel, 1990].

'largeness principles' in connection with solving the continuum hypothesis. He suggests that they might be added in a cumulative fashion, with some 'domain of propositions in which you are interested' becoming decided at some level of this cumulative procedure:

> Let us consider, e.g, the concept of demonstrability. It is well known that, in whichever way you make it precise by means of a formalism, the contemplation of this very formalism gives rise to new axioms which are exactly as evident and justified as those with which you started, and that this process of extension can be iterated into the transfinite. So there cannot exist any formalism which would embrace all these steps; but this does not exclude that all these steps (*or at least all of them which give something new for the domain of propositions in which you are interested*) could be described and collected together in some non-constructive way. In set theory, e.g., the successive extensions can most conveniently be represented by stronger and stronger axioms of infinity. It is certainly impossible to give a combinatorial and decidable characterization of what an axiom of infinity is; but there might exist, e.g., a characterization of the following sort: An axiom of infinity is a proposition which has *a certain (decidable) formal structure* and which in addition is true. Such a concept of demonstrability might have the required closure property, i.e., the following could be true: Any proof for a set-theoretic theorem in the next higher system above set theory (i.e., any proof involving the concept of truth which I just used) is replaceable by a proof from such an axiom of infinity. It is not impossible that for such a concept of demonstrability some completeness theorem would hold which would say that every proposition expressible from set theory is decidable from the present axioms plus some true assertion about the largeness of the universe of all sets. (emphasis ours)

For a further analysis of Gödel's views on extensions of the Zermelo-Fraenkel axioms, also in relation to the present developments in set theory, see [Kennedy and van Atten, 2004].

1.4 Speed-up Theorems

Gödel submitted an abstract in 1936 for what came to be known as the 'speed-up' theorem, entitled 'On the length of proofs.'[126] This result says that while some sentences of arithmetic are true but unprovable, there are other sentences which are provable but even the shortest proof is longer than any bound given in advance. More exactly:

THEOREM 9. *Given any recursive function f there are provable sentences ϕ of arithmetic such that the shortest proof is greater than $f(\ulcorner\phi\urcorner)$ in length.*

[126] See [Gödel, 1986].

The Speed-up Theorem is the result of contemplating and elaborating the proof of the incompleteness theorem. It applies the fixed-point technique to the concept of unprovability by a short proof, as opposed to the original idea of applying the fixed-point theorem to mere unprovability. The proof has very much the same flavor as the proof of the incompleteness theorem. Interestingly, it dates from the same year as the construction, due to Rosser, that eliminates the use of ω-consistency in the first Incompleteness Theorem; like the Speed-up Theorem of Gödel, Rosser's construction exploits the issue of short and long proofs.

Gödel never submitted a proof for the Speed-up Theorem. Over the years several related proofs were published, but the first full proof of Gödel's original result was given only in 1994 by Sam Buss in his 'On Gödel's theorems on lengths of proofs I: Number of lines and speedups for arithmetic.'[127] Buss also gives a second proof of the theorem which avoids self-reference, following a technique of Statman.

Gödel measures the length of proofs by the number of formulas. There are also other possibilities, like the number of symbols in the proof. The case of the Speed-up Theorem where the length of proof is measured by the number of symbols, was proved by Mostowski in 1952.[128] Though both measures may be equally natural candidates for measuring the length of a proof, proving the theorem for length measured by the number of symbols avoids a technical complication introduced by the other measure: there are only finitely many proofs with a given number of symbols, whereas there are infinitely many proofs with a given number of formulas.

Gödel states the Speed-up Theorem differently from the above. Let S_n be the system of logic of the n-th order, the variables of the first level being thought of as ranging over natural numbers. In this setting, variables of the second level range over sets of natural numbers and so on. Gödel's formulation is:

THEOREM 10. *Let n be a natural number > 0. If f is a computable function, then there are infinitely many formulas A, provable in S_n, such that if k is the length of the shortest proof of A in S_n and l is the length of the shortest proof of A in S_{n+1}, then $k > f(l)$.*

We note that the theorem depends on the fact that the notion of computable function is absolute, i.e. independent of n. We now sketch the proof measuring length as the number of symbols:

To prove Theorem 9, let f be total recursive function. By Gödel's Fixed Point theorem there is a formula $\phi(n)$ stating '$\phi(n)$ has no proof in PA shorter than $f(n)$'. This is tenable if the length is measured by number of symbols, because we only need to search through finitely many proofs shorter than $f(n)$.

Note that $\phi(n)$ is true for all n, for if $\phi(n)$ were false, then there would be a short proof of $\phi(n)$, and hence by soundness $\phi(n)$ would be true, a contradiction: $\phi(n)$ would both true and false. This can be formalized in PA and thus we get

[127] See [Buss, 1994].

[128] See [Mostowski, 1982]. For proofs of similar results see [Ehrenfeucht and Mycielski, 1971], and [Parikh, 1971].

the result that for each n the sentence $\phi(n)$ is provable in PA. Since $\phi(n)$ is true for all n, it cannot have a proof in PA which is shorter than $f(n)$.

For Theorem 10, suppose we have a higher type system in which we can prove $\forall x\phi(x)$. This proof is of constant length. Thus each $\phi(n)$ is derivable from this universal statement by one application of the logical rule $\forall x\phi(x) \rightarrow \phi(t)$. Thus $\phi(n)$ has in that system for all n a short proof. What kind of stronger system can we have in which $\forall x\phi(x)$ is provable? We may consider second order logic in which we can define a predicate $N(x)$ for the set of natural numbers and furthermore can prove of a new predicate symbol $Tr(x)$ that it satisfies the inductive clauses of the truth definition of first order formulas of arithmetic, relativized to N. Then the stronger system can prove that provable first order sentences of arithmetic satisfy the predicate Tr. By the above argument, we can prove in the stronger system that $\forall x\phi(x)$ satisfies Tr. Then by adding a few lines we can prove each $\phi(n)$ satisfies Tr. Because of the nature of $\phi(n)$, this implies the stronger system has a (short) proof of $\phi(n)$. An alternative system is Peano's axioms PA in an extended language where we have a new predicate symbol Tr and axioms stating that the predicate Tr codes the satisfaction relation for all sentences of the vocabulary not containing Tr.

The proof we have outlined is sensitive to the particular concept we use for the length of a proof. Another possibility, and the one that Gödel has in mind as we saw, is the number of formulas in the proof. Buss proves the theorem in either case, so both cases are resolved.

An alternative way of getting speed-up is to take any unprovable sentence as a new axiom. The new axiom shortens dramatically proofs of some sentences. Those sentences used to be provable anyway but are now provable with a much shorter proof.[129]

THEOREM 11. *Suppose $P \nvdash \phi$. Let f be a recursive function. Then there is a sentence ψ such that $P \vdash \psi$, and a proof p of ψ in $P \cup \{\phi\}$, such that there is no proof of ψ from P which is shorter (in symbols) than $f(n)$ where n is the length of p.*

Proof. Suppose there is no such ψ. Then for all ψ such that $P \vdash \psi$, let $h(\psi)$ be some proof of $\phi \vee \psi$ in $P \cup \{\phi\}$. Now for all ψ, $P \cup \{\neg\phi\} \vdash \psi$ if and only if $P \vdash \phi \vee \psi$ if and only if there is a proof of $\phi \vee \psi$ from P of length $\leq f(h(\phi \vee \psi))$. But this means that $P \cup \{\neg\phi\}$ is decidable, contrary to the essential undecidability of P. ■

2 GÖDEL'S RESULTS IN INTUITIONISTIC LOGIC AND ARITHMETIC

Abbreviations Following [Troelstra and van Dalen, 1988], we will use these abbreviations:

[129] See also [Statman, 1981].

IPC	intuitionistic propositional logic
IQC	intuitionistic predicate logic
CPC	classical propositional logic
CQC	classical predicate logic
PA	Peano Arithmetic
HA	Heyting Arithmetic (PA with IQC instead of CQC as the underlying logic)
MP	Markov's Principle

Gödel's results The results on intuitionistic systems obtained by Gödel and published by him are, in chronological order of publication:

1932	IPC is not a finitely-valued logic
1933	PA is translatable into HA
1933	IPC is translatable into classical S4
1958	HA is interpretable in terms of computable functionals of finite type

A result due to him and published by Kreisel is

1957	Semantical completeness of IQC implies validity of MP (published in 1962)

Background: Heyting's Proof Interpretation A characteristic feature of the published results is that none of them is concerned with the intuitionists' intended interpretation of IPC, IQC, or HA (except in the sense of wishing to avoid it). The intended interpretation of IQC was given by Heyting and is known as 'the Proof Interpretation', where proofs are understood as mental objects:[130]

1. A proof of $A \wedge B$ is given by presenting a proof of A and a proof of B.

2. A proof of $A \vee B$ is given by presenting either a proof of A or a proof of B.

3. A proof of $A \rightarrow B$ is a construction which permits us to transform any proof of A into a proof of B.

4. Absurdity $\perp$ (contradiction) has no proof; $\neg A$ is defined as $A \rightarrow \perp$, so a proof of $\neg A$ is a construction which transforms any hypothetical proof of A into a proof of a contradiction.

5. A proof of $\forall x A(x)$ is a construction which transforms any d in the domain into a proof of $A(d)$.

6. A proof of $\exists A(x)$ is given by presenting a specific d from the domain and a proof of $A(d)$.

[130] [Heyting, 1930a], [Heyting, 1956, section 7.1.1], [Troelstra and van Dalen, 1988, 1.3.1]

Discussion

IPC is not a finitely-valued logic Answering a question posed by Hahn[131]—and apparently without knowing Becker's [1927]; see below—, in his note 1932 Gödel demonstrates two propositions:

1. IPC cannot be viewed as a logic with finitely many truth values.

and

2. There is an infinite, monotonically decreasing sequence of logics that are weaker than CPC but stronger than IPC.

The question to which the first result (which is a consequence of the second[132]) provides an answer had been topical for some years. In 1927, Barzin and Errera had proposed a conception of IPC according to which it is three-valued; moreover, they argued that, given the principles they considered valid for propositions having the third truth value, IPC is inconsistent.[133] Also in 1927, Oskar Becker had made the contrary conjecture that in intuitionistic logic there is no 'n-tum non datur'.[134] That Barzin and Errera's argument was incorrect was quickly pointed out by Church.[135]

Gödel's argument settled the general question in a definitive and completely formal manner. Like his incompleteness theorem, this was a limitative result; this led Heyting to remark, in a letter to Gödel dated November 24 and 26, 1932, 'It is as if you had a malicious pleasure in showing the purposelessness of others' investigations'—adding that 'In the sense of economy of thought this work is certainly useful, and in addition to that comes the particular beauty of your short proof'.[136]

PA is translatable into HA In [1933e], reporting on the meeting of Menger's colloquium of June 28, 1932, Gödel presents a translation $'$ from PA into HA such that

$$\text{If PA} \vdash A \text{ then HA} \vdash A'$$

(Gödel does not actually use PA but the system of Herbrand.[137]) The converse also holds, as $\text{PA} \vdash A \leftrightarrow A'$. The translation $'$ Gödel defined as follows:

[131] See [Gödel, 1933] and [Gödel, 1933n].
[132] [Gödel, 1986, p.223]
[133] [Barzin and Errera, 1927]
[134] [Becker, 1927, pp.775–777]
[135] [Church, 1928]. For a discussion of the whole episode, see [Hesseling, 2003, pp.260–266].
[136] [Gödel, 2003b, p.67]
[137] [Herbrand, 1931]

$$\begin{array}{rcl} (\neg A)' & := & \neg A \text{ (A atomic)} \\ (A \to B)' & := & \neg(A' \wedge \neg B') \\ (A \vee B)' & := & \neg(\neg A' \wedge \neg B') \\ (A \wedge B)' & := & A' \wedge B' \\ \forall x A(x)' & := & \forall x A'(x) \end{array}$$

The translation is also known as 'the negative translation' as its range is in the negative fragment of intuitionistic logic (i.e., the fragment not containing $\exists$ or $\vee$, and in which atomic formulas only appear with a negation sign in front of them). Gödel's actual clause for $\forall$ is a little more complicated, as HA as we know it is really an extract from Heyting's original system of wider scope, which had only variables for arbitrary objects and none specifically for numbers. No clause for $\exists$ is given by Gödel, but classically it can of course be defined in terms of $\forall$ and $\neg$.

Hesseling[138] notes that, first, the translation Gödel proposes is incomplete, for it omits atomic formulas; and second, that this is not resolved in the introductory note in the Collected Works, for it contains the misprint '$P' := P$ (P atomic)' instead of '$(\neg P)' := \neg P$ (P atomic)'. Two solutions are proposed by Hesseling: to limit the domain of the translation and consider only negated atomic formulas, or to add the clause

$$A' \quad := \quad \neg\neg A \text{ (A atomic)}$$

This clause is motivated by a wish for generality with an eye on applications of $'$ also outside arithmetic, but in the specific (and original) case of arithmetic, one could also leave atomic formulas unchanged, because $\mathrm{HA} \vdash \neg\neg(a = b) \leftrightarrow a = b$. That equivalence can be used to show that for A not containing $\vee$ or $\exists$, one even has

$$\mathrm{PA} \vdash A \text{ if and only if } \mathrm{HA} \vdash A$$

i.e., without need for translation.

If we denote derivability in IQC by $\vdash_i$ and derivability in CQC by $\vdash_c$, and define the translation of a set of formulas by $\Gamma' = \{B' | B \in \Gamma\}$, then the translation $'$ serves to show

$$\Gamma \vdash_c A \text{ if and only if } \Gamma' \vdash_i A'$$

This way, Gödel's note extends work by Kolmogorov and Glivenko;[139] only the latter paper was known to Gödel at the time.[140]

Gödel goes further by turning to arithmetic. He shows that if A is provable in PA, then A' is provable in HA, and observes that this provides an intuitionistic consistency proof of classical arithmetic and number theory. This is because $(0 \neq 0)'$ is $0 \neq 0$.

[138] [Hesseling, 2003, pp.287–288]

[139] [Kolmogorov, 1925], [Glivenko, 1929]

[140] See [van Atten, 2004b, p.426] for some remarks on the historical relation between these two Russian papers.

Gödel's result had been obtained independently by Gentzen (with a contribution by Bernays); Gentzen leaves atomic formulas unchanged, and preserves implication by defining $(A \to B)' := A' \to B'$. Upon learning of Gödel's forthcoming paper, Gentzen stopped the publication of his own, of which galley proofs had already been made. In a letter to Heyting of May 16, 1933, Gödel writes that in Göttingen his result should have been known since June 1932, because Veblen had been present at his presentation at Menger's colloquium and then travelled to Göttingen;[141] Heyting, who had received Gentzen's manuscript before (an offprint of) Gödel's, replies on August 24, 1933, 'I must assume that the report about your lecture had not sufficiently got through to Göttingen after all'.[142]

Different aspects of the Gödel-Gentzen translation may be selected for emphasis: it shows that, whatever advantages HA may (in a foundational debate) be claimed to have over PA, these must reside in its interpretation and not in its safety; it also shows that HA can be used as a means to establish consistency of PA. No doubt intuitionists will pay more attention to the first, classical mathematicians to the second aspect.

Kreisel has noticed the contrast between Gödel's and Gentzen's views on the first aspect: Gödel takes his translation to show that intuitionistic logic involves only 'eine etwas abweichende Interpretation',[143] whereas Gentzen stressed that the sense in which intuitionistic logic is part of classical logic has only formal significance, thereby drawing attention to the (semantical) richness of intuitionistic logic.[144] Later Gödel became, in Kreisel's words, 'supersensitive about differences in meaning'.[145] (In analysis the situation is different even purely formally, as there the intuitionistic formal systems are not proper parts of classical ones.)

Regarding the second aspect, Bernays wrote that it convinced the Hilbert school that, 'contrary to the prevailing views at the time', intuitionism was stronger than the finitary standpoint, and that new methods of reasoning could be looked for that are non-finitary but all the same 'of a constructive character'.[146] Gentzen would soon give his consistency proof using transfinite induction up to ϵ_0, and later Gödel would devise another constructive but non-finitary method in his Dialectica Interpretation (see below).

IPC is translatable into classical S4 The note [1933f] communicates a result that Gödel had presented in Menger's colloquium, at an unknown date in the academic year 1931–1932:

If IPC $\vdash A$ then S4 $\vdash A'$

[141] [Gödel, 2003b, p.71]
[142] [Gödel, 2003b, p.71]
[143] [Gödel, 1933e, p.294]
[144] [Kreisel, 1971, p.257]
[145] [Kreisel, 1987, p.82]
[146] [Bernays, 1967, p.502]; Bernays had made a remark to this effect in print already [Bernays, 1935, p.271].

where $'$ is a translation from IPC to S4 defined as follows (the connectives in S4 are to be taken classically):

$$\begin{array}{rcl} (\neg A)' & := & \neg\Box A' \\ (A \to B)' & := & \Box A' \to \Box B' \\ (A \vee B)' & := & \Box A' \vee \Box B' \\ (A \wedge B)' & := & A' \wedge B' \end{array}$$

Gödel conjectured that the implication also holds in the other direction, and planned to work on this during his stay in Princeton in 1933–1934;[147] eventually, the conjecture was verified by McKinsey and Tarski.[148]

As is characteristic for Gödel's work on intuitionism, this translation provided him with a means to study IPC while steering clear of its intuitionistic interpretation;[149] equally characteristic is the formal nature of the alternative interpretation, and in this respect, Gödel himself notes, his result differs from Kolmogorov's paper of 1932.[150] A forerunner that is closer in spirit to Gödel's translation is Becker's 'Zur Logik der Modalitäten',[151] which he also cites. Becker discussed Heyting's formalization of intuitionistic logic from 1930,[152] and attempted (but failed) to clarify the differences between classical and intuitionistic logic by interpreting the latter in a system of modal logic that anticipated S4.[153] In particular, this system included what is sometimes called (also by Gödel in this note) 'Becker's axiom', $\Box A \to \Box\Box A$. As the axiom '4' it was then adopted by Lewis and Langford in 1932,[154] when presenting the extension of Lewis' original system of 1918 that they baptized S4. Gödel however does not actually describe his system as a (new) axiomatization of S4, but as an equivalent of 'Lewis' system of strict implication' when the latter is supplemented by Becker's axiom.

A translation of IPC into a modal logic (anticipating S4) had been given before Becker and Gödel, who evidently were unaware of this fact, by the Russian logician I.E. Orlov.[155]

As pointed out by Troelstra in his introduction to Gödel's translation,[156] at the time Gödel certainly knew the content of [Heyting, 1931]. He had attended the Königsberg conference, at which Heyting had presented that text, and published a review of the printed version.[157] In that text, Heyting identifies the notions of being intuitionistically true and being intuitionistically provable. Heyting also explicitly discusses a provability operator (but does not go on to present a modal logic for it). Hence, as Troelstra concludes, it may well have been Heyting's lecture

[147] See his letter to von Neumann, March 14, 1933 [Gödel, 2003b, pp.346–347].
[148] [McKinsey and Tarski, 1948]
[149] See also Gödel's remarks on this in the Lecture at Zilsel's [Gödel, 1995, pp.100–101].
[150] [Kolmogorov, 1932]
[151] [Becker, 1930]
[152] [Heyting, 1930a],[Heyting, 1930b]
[153] See for a discussion of Becker's work [Peckhaus, 2002].
[154] [Lewis and Langford, 1932]
[155] See [Došen, 1992], [Bazhanov, 2003].
[156] [Gödel, 1986, p.299]
[157] [Gödel, 1932f]

that led Gödel to develop this interpretation (in particular, one may add, if at the same time Gödel was aware of the problems left in [Becker, 1930]). In a letter of January 12, 1931, von Neumann had suggested a kind of provability logic to Gödel,[158] where 'provability' is taken in the sense of Gödel's predicate 'Bew' in [Gödel, 1931]. As a matter of psychology, this suggestion may or may not have been instrumental in making Gödel's discovery of his interpretation of IPC; but, for the reason given in the next paragraph, what von Neumann suggested is not at all what Gödel has in mind here.

Gödel concludes his note by observing that his provability operator cannot be interpreted by 'being provable in a certain formal system': $\Box(\Box A \to A)$ is a theorem of this provability logic, but for no formal system that satisfies the appropriate conditions one has $\Box(0 \neq 0) \to 0 \neq 0$, as its immediate consequence $\neg\Box(0 \neq 0)$ would contradict the second incompleteness theorem. Thus, the desideratum of a truth-function style semantics for this provability logic based on S4 remained; it was first met by Artemov.[159]

The interpretation of IPC in terms of a provability operator will differ from the intended intuitionistic one if one wishes to leave open the possibility that a proposition A is provable without this being known. For in that case, to say that A is provable is to say that there exists a proof of it in a sense that transcends the realm of the mathematician's observations on his own constructional activity; but according to intuitionists, mathematics should not go beyond that realm. On their conception, as Heyting says, 'the possibility of a construction can be proved only by its actual execution' (which is not incompatible with introducing certain idealisations of the notion of actual execution). But in that case, Heyting points out, an explicit provability operator is redundant.[160]

Semantical completeness of IQC implies validity of MP Although from a principled intuitionistic point of view Heyting's Proof Interpretation provides a perfectly adequate semantics, several alternatives have been proposed that lend themselves more easily to formal study, with the aim of arriving at an intuitionistic counterpart to classical model theory. Examples are Beth models and Kripke models, with the corresponding notions of validity. Another option is to define intuitionistic validity analogously to the set-theoretical notion of classical validity; call this 'structural validity'. Let A be a formula of IQC containing the predicate symbols $P_1 \dots P_k$. Then $A(P_1 \dots P_k)$ is valid if for all intuitionistically meaningful domains (species) $D*$ and all relations $P_1^* \dots P_k^*$ defined on $D*$, $A(P_1^* \dots P_k^*)$ holds intuitionistically, where all quantifiers in A are relativized to $D*$ and P_i is interpreted by P_i^*; notation $\vDash A$. The intended interpretation of the connectives is presupposed. IQC is said to be (semantically) complete for a formula A if $\vDash A$ implies $\vdash A$; it is said to be (semantically) complete if it is (semantically) complete

[158] [Gödel, 2003b, pp.340–345]
[159] [Artemov, 2001]
[160] [Heyting, 1931, p.60]

for every A.[161]

Gödel demonstrated that completeness of IQC with respect to structural validity constructively implies MP in the form

$$\forall\alpha_{0,1}\neg\neg\exists x A(\alpha, x) \rightarrow \forall\alpha_{0,1}\exists x A(\alpha, x) \qquad (\mathrm{MP}_\alpha)$$

and, as a corollary, also in the form

$$\neg\neg\exists x A(x) \rightarrow \exists x A(x) \qquad (\mathrm{MP}_{\mathrm{PR}})$$

where in both A is primitive recursive, and $\alpha_{0,1}$ ranges over choice sequences (not: lawless sequences) chosen from $\{0, 1\}$ (the restriction to 0 and 1 is convenient but not essential).

A proof of Gödel's result was published in 1962 by Kreisel, who specifies that Gödel related his arguments to him in conversation in 1957;[162] Kreisel simplified Gödel's own proof, using suggestions by Kleene, and investigated it further. Gödel's argument combines theorem X of his incompleteness paper[163] with the constructive proof of Herbrand's Theorem.[164]

On the Proof Interpretation, $\mathrm{MP}_{\mathrm{PR}}$ is not acceptable. $\neg\neg\exists x A(x)$ means that it is contradictory to assume that one cannot prove $A(n)$ for some n; but the contradictoriness of that assumption in general may be too little information to enable one to construct an n such that $A(n)$, in which case such an n cannot, in the intuitionistic sense, be said to exist.[165]

Gödel's argument, which also holds for the notions of validity defined by Beth and Kripke semantics, seems to show that the intuitionist cannot hope ever to establish completeness of IQC. (It is a different question whether an intuitionist would attach the same significance to completeness of a formal system as birds of different feathers would.) Is this indeed the case? We note three different views.

1. The most widely held view seems to be that principles such as $\mathrm{MP}_{\mathrm{PR}}$ are not intuitionistically valid, that the notions of validity for which Gödel's argument goes through are in some sense correct, and that therefore that argument rules out an intuitionistic argument for completeness of IQC.[166]

2. McCarty does not question these notions of validity either, but attempts to argue—specifically, against Dummett—that $\mathrm{MP}_{\mathrm{PR}}$ is intuitionistically acceptable after all, not as a principle of logic but as a principle of arithmetic.[167]

[161] For discussion of structural validity in an intuitionistic context, see [Troelstra, 1977, pp.102–104], [Dummett, 2000, pp.154–159], as well as the references in footnote 168.
[162] [Kreisel, 1962], in particular p.142. See also the abstract [Kreisel, 1958].
[163] [Gödel, 1931, pp.186–187]
[164] See also [Troelstra and van Dalen, 1988, ch.13, section 3] and [Dummett, 2000, pp.177–180]. A different kind of proof is in [McCarty, 1994].
[165] See [Dummett, 2000, pp.174–175] for an elaboration of this argument.
[166] [Dummett, 2000, pp.14, 174, 186]
[167] [McCarty, 1994, pp.105–106]

3. Some intuitionists reject $\mathrm{MP}_{\mathrm{PR}}$ and also question the intuitionistic appropriateness of the notions of structural, Beth and Kripke validity.[168] In particular, they have pointed out that Gödel's argument depends on the way in which negation is treated in these three types of model (i.e., negation as absence of models), and have argued that an intuitionist is not committed to treating it that way. For example, Veldman showed how, by generalizing the notion of a Kripke model a little (by allowing $\bot$ actually to hold at some nodes, while strengthening the condition on validity of negated sentences), a constructive completeness proof for IQC can be given that escapes Gödel's argument.[169] De Swart showed that a similar result can be obtained by generalizing Beth models.[170] But Veldman also indicated that, from an intuitionistic point of view, the Henkin-style semantics he arrived at is a formal artifice and does not much to clarify the notion of an intuitionistically true sentence.[171] De Swart then went on to propose yet another notion of model, argued that this one does yield an intuitionistically natural semantics, gave a completeness proof relative to this semantics, and pointed out that here too Gödel's argument does not go through.[172] This semantics seems natural indeed, but this remains a matter of some debate.

HA is interpretable in terms of computable functionals of finite type In his paper [1958] (written in honour of Bernays' 70th birthday), Gödel presents a system T (a theory of finite types) and a translation $'$ of formulas F from HA to T such that

(i) F' is of the form $\exists y \forall z A(y, z, x)$

(ii) If $\mathrm{HA} \vdash F$, then in T functions Q can be defined such that $\mathrm{T} \vdash A(Q(x), z, x)$ (This is not actually proved in the paper.)

Here x refers to all free variables of F, y and z are finite sequences of variables of any type, and $A(y, z, x)$ contains no quantifiers and no variables that are not in x, y or z; we will remark on the idea behind this particular form of F' after defining $'$ and giving the intended interpretation of T.

The translation F' of a formula F is defined by induction on the number k of logical operators in F ($s \ldots z$ and $V \ldots Z$ stand for finite (possibly empty) sequences of, respectively, arbitrary type and higher type; in particular, x and u denote the sequences of free variables in A and B):

For $k = 0$ (i.e., F atomic), $F' := F$

[168] [Lopez-Escobar and Veldman, 1975, p.201], [de Swart, 1977a, p.565]
[169] [Veldman, 1976]; see also [Lopez-Escobar and Veldman, 1975].
[170] [de Swart, 1977a]
[171] [Veldman, 1976, p.159]
[172] [de Swart, 1977b]

For the induction step, suppose $F' = \exists y \forall z A(y,z,x)$ and $G' = \exists v \forall w B(v,w,u)$; then

$$\begin{array}{rcl}
(F \wedge G)' & := & \exists yv \forall zw[A(y,z,x) \wedge B(v,w,u)] \\
(F \vee G)' & := & \exists yvt \forall zw[t = 0 \wedge A(y,z,x) \vee t = 1 \wedge B(v,w,u)] \\
(F \rightarrow G)' & := & \exists V Z \forall yw[A(y, Z(yw), x) \rightarrow B(V(y), w, u)] \\
(\exists sF)' & := & \exists sy \forall z A(y,z,x) \\
(\forall sF)' & := & \exists Y \forall sz A(Y(s), z, x) \\
(\neg F)' & := & \exists \overline{Z} \forall y \neg A(y, \overline{Z}(y), x)
\end{array}$$

Clearly, PA can be shown to be consistent by showing that T is, because one already has a consistency proof of PA relative to HA (see above). To establish consistency of T, Gödel takes the computable functionals of finite type as a model. In the revised version [1972], first published in the Collected Works, two further models are proposed, the hereditarily recursive operations and a term model, with which we will not be concerned. The computable functionals of finite type are a generalization of Turing's computable functions. Gödel defines them as follows:[173]

1. The computable functionals of type 0 are the natural numbers
2. If $t_0, t_1, \ldots, t_k$ are previously defined types ($k \geq 1$), then also $(t_0, t_1, \ldots, t_k)$ is a type, namely, the type of computable functionals that assign to every k-tuple of computable functionals of types $t_1, \ldots, t_k$ a computable functional of type t_0. This assignment is to take place by an operation that is always performable and is constructively recognizable as such.

The general notion of computable functional of finite type is then arrived at by considering the type t as a variable. This notion is to be taken as primitive, and is, as Gödel says, abstract in that it does not 'involve properties or relations of concrete objects (for example, combinations of signs)'. In Hilbert's conception of finitary mathematics, evidence which is abstract in this sense is not admitted. The properties involved in abstract notions are 'essentially of higher order' and pertain for example to meanings of terms. In 1958, Gödel describes this by saying that abstract notions relate to 'mental constructs'; in 1972, he has replaced this term by 'thought structures or thought contents'.[174] Even though further on in the sentence he refers to 'mental objects' again, arguably the replacement made fits an objectivist conception of mathematics somewhat better. In any case the reference to the 'mental' both in 1958 and in 1972 should not be taken in the sense in which Brouwer holds that mathematics essentially is nothing but a mental construction (on this issue, see also Bernays' distinction between 'reserved intuitionism' and 'far-reaching intuitionism' in his letter to Gödel of March 16, 1972.[175]) We will have occasion to come back to this below.

To see the idea behind the definition of the clauses of the translation, consider the constructive meaning of $\exists x A(x) \rightarrow \exists u B(u)$: one has a construction that, given

[173] [Gödel, 1990, p.245]
[174] [Gödel, 1990, p.241] and [Gödel, 1990, p.273]
[175] [Gödel, 2003a, pp.294–295]

an object a such that $A(a)$, yields an object b such that $B(b)$. Assuming that A and B are atomic, the translation of this formula is $\exists U \forall x(A(x) \to B(U(x)))$. If we interpret U as a computable functional, this translation impeccably meets the constructive requirement. The higher-order functions reflect the interplay between, on the one hand, the connectives and, on the other, the first-order quantifiers as constructively construed, and thereby serve to remove quantifiers from the connected statements. In this particular sense, the Dialectica Interpretation is a form of quantifier elimination. As statements without quantifiers are decidable, the connectives between them become simple truth functions.

The point of the paper is an epistemological one, and fits Gödel's general interest in the investigation of the limits (approached from both sides, so to speak) of formal methods in capturing intuitive concepts.[176] Soon after publishing his two incompleteness theorems, Gödel had come to accept the view that the second demonstrates that Hilbert's original goal of establishing the consistency of PA by finitary means is unattainable.[177] But, as Bernays has written (see above), the proofs by Gödel and Gentzen that PA is translatable into HA inspired hope that the consistency of PA could be established by means that, although not finitary, are still constructive. With his Dialectica Interpretation, Gödel introduces the abstract notion of his computable functionals as such a means, as an alternative to Heyting's Proof Interpretation as well as to Gentzen's induction up to ϵ_0. The odd impression that a consistency proof of arithmetic in terms of a notion that includes the natural numbers themselves as a special case may make, is removed by realizing that (interpreted) PA and HA appeal to either non-constructive or abstract notions via their logic, which is after all also part of these systems: through the use of computable functionals of finite type, logical complexity is traded in for ontological complexity.

In the following, we will be specifically concerned with the comparison of the Dialectica Interpretation with Heyting's Proof Interpretation.[178]

From Gödel's 1958 paper it is clear (with a minimum of reading between the lines) that he considered his interpretation to be superior to Heyting's, the notion of computable functional being more evident than the intuitionistic notion of proof. In the revision from 1972 this aspect is made more explicit, in particular in footnote h:[179]

> The interpretation of intuitionistic logic, in terms of computable functions, in no way presupposes Heyting's and ... moreover, it is constructive and evident in a higher degree than Heyting's. For it is exactly the elimination of such vast generalities as 'any proof' which makes for greater evidence and constructivity.

[176] Yourgrau [2005, p.182] calls this 'Gödel's Program'.

[177] Compare [Gödel, 1931, p.195] with [Gödel, *1933o, p.52 (Gödel's page 26)]. See also [Mancosu, 1999].

[178] For Gödel's view on Gentzen's approach, see e.g. footnotes 3, 4 and f in the version of 1972 [1972, pp.273–274], as well as sections VI and VII of his lecture at Zilsel's [*1938a, pp.102–113].

[179] [Gödel, 1990, p.276]

This is one of the differences between the two versions that reflect, as Gödel wrote to Bernays in 1968 while working on the revision, how in 1958 he 'placed no particular value on the philosophical matters' because he was 'mainly concerned with the mathematical result,' while 'now it is the other way around'.[180] It is quite likely that the increase in sensitivity to questions of evidence that Gödel displays in the 1972 version is due to his wide reading in phenomenology that he had begun shortly after completing the original version from 1958.

Gödel's qualm with Heyting's Proof Interpretation, as voiced in his Cambridge lecture in 1933, is that the clause for negation (more generally, the clause for implication) involves the notion of arbitrary intuitionistic proof in an essential way, and that this notion is too indeterminate. It does not comply with a condition he (at the time) poses on constructivity:[181]

> Heyting's axioms concerning absurdity and similar notions ... violate the principle, which I stated before, that the word 'any' can be applied only to those totalities for which we have a finite procedure for generating all their elements ... The totality of all possible proofs certainly does not possess this character, and nevertheless the word 'any' is applied to this totality in Heyting's axioms ... Totalities whose elements cannot be generated by a well-defined procedure are in some sense vague and indefinite as to their borders. And this objection applies particularly to the totality of intuitionistic proof because of the vagueness of the notion of constructivity.

An intuitionist might reply that we recognize a proof when we see one (Kreisel); the clause for implication (and hence that for negation) is not to be taken as quantifying over an actual totality of intuitionistic proofs, but rather as a schema: 'I have a construction that, whenever you give me something recognizable as a proof of the antecedent of $A \to B$, will transform this into a proof of the consequent.' This can be expected to work because in proofs of implications usually nothing more is assumed about a proof of the antecedent than that it indeed is one (see also the remark on impredicativity, below). The prime example of an intuitionistic theorem that goes beyond that assumption is Brouwer's proof of the bar theorem.[182] This (classically trivial, but constructively remarkable) theorem basically says that, if a tree contains a subset of nodes such that every path through the tree meets it (a 'bar'), then there is a well-ordered subtree that contains a bar for the whole tree. Brouwer's extraction of additional information from the hypothesis that we have a proof of the antecedent (i.e., that we have a proof that the tree contains a bar) is based on his analysis, perfectly in accord with his general conception of mathematics, of proofs in terms of mental structures. That analysis enables him to formulate a necessary condition on proofs of the antecedent, namely, that they admit of being put into a canonical form;[183] on the basis of that form, a proof

[180] [Gödel, 2003a, pp.260–261]
[181] [Gödel, *1933o, p.53]
[182] [Brouwer, 1924], [Brouwer, 1927]
[183] [Brouwer, 1927, p.64]

of the consequent is obtained. Thus, it is by devising a transcendental argument that Brouwer avoids issues of impredicativity here (see [Sundholm and van Atten, in print]). To Gödel however, relying on properties of mental structures must have amounted to introducing a non-mathematical element. More on this below.

Here we note that Gödel's dissatisfaction with the Proof Interpretation was at the root of the development of the Dialectica Interpretation. That development can roughly be traced as follows:[184]

1933: Lecture 'The present situation in the foundations of mathematics'.[185] Gödel formulates his objection to the Proof Interpretation, which interpretation he considers 'of doubtful value' in proving the consistency of PA. He expresses the hope that 'in the future one may find other and more satisfactory methods of construction' that go beyond finitary reasoning.

1938: 'Lecture at Zilsel's'.[186] Gödel suggests that higher types may be used to arrive at a consistency proof of PA. However, there is no indication of a method, nor of a (possible) interpretation of HA in higher types. However, Gödel does add that with a fixed finite number of higher types, consistency cannot be shown.

1941: Lecture 'In what sense is intuitionistic logic constructive?'.[187] This time, Gödel phrases his objection to the Proof Interpretation by saying that the clause for implication requires that 'the notion of derivation or of proof must be taken in its intuitive meaning as something directly given by intuition, without any further explanation being necessary. This notion of an intuitionistically correct proof or constructive proof lacks the desirable precision.'[188] Then Gödel presents, in an informal manner, an interpretation of HA in a system of higher types Σ, and explains the motivation behind the interpretation and behind its clauses. Gödel introduces the schemata of Explicit Definition and Recursive Definition to define the constructive operations referred to in the definition of the computable functionals. Some applications are given, for example, if $\mathrm{HA} \vdash \exists x \phi(x)$, then from the corresponding proof of $\exists x \phi'(x)$ in Σ one obtains a witness; this is then used to show that if $\mathrm{PA} \vdash \exists \phi(x)$, where ϕ is decidable (contains no quantifiers), then there is a t such that $\Sigma \vdash \phi(t)$. Finally, Gödel suggests the interpretation may be extended from arithmetic to analysis.

1958: First publication of the interpretation.[189] The system T is very similar to the system Σ of 1941, but Gödel no longer tries to define the notion of constructive operation and takes it as primitive instead. Gödel now claims superiority over the Proof Interpretation by saying, as we saw above, that his own interpretation is 'constructive and evident in a higher degree' because of its 'elimination of such vast generalities as "any proof" which makes for greater evidence and constructivity'.

Philosophically, this difference between the approaches in 1941 and 1958—trying to find suitable schemata for defining the constructive operations used in the definition of computable function versus taking the notion as primitive—marks a sea change. In footnote 6, Gödel writes in defence of the latter approach:[190]

[184] For extensive discussion, see Troelstra's introductions to [Gödel, *1941] and [Gödel, 1958; Gödel, 1972] in [Gödel, 1990], and 'Gödel's Dialectica Interpretation and its Two-Way Stretch' in [Feferman, 1998].
[185] [Gödel, *1933o]
[186] [Gödel, *1938a]
[187] [Gödel, *1941]
[188] [Gödel, *1941, p.190]
[189] Kreisel, who had learned of the essentials of the interpretation from Gödel himself in 1955, had made it public (again) in lectures at Cornell and at Amsterdam in 1957 [Kreisel, 1987, pp.104, 109].
[190] [Gödel, 1990, p.245n.6]

> As is well known, A.M. Turing, using the notion of a computing machine, gave a definition of the notion of computable function of the first order. But, had this notion not already been intelligible, the question whether Turing's definition is adequate would be meaningless.

Mutatis mutandis, he could have written this about his former self. In Gödel's view as expressed in 1958, what Turing did was to define (and in that sense see sharper) an objective concept that we had been perceiving all along, albeit less sharply.[191] Similarly, Gödel holds, there is an objective concept 'computable functional', that we may not yet (and possibly never) be able to make completely explicit, but that at the same time we see enough of to determine some of its properties. As he wrote in the preceding footnote:[192]

> One may doubt whether we have a sufficiently clear idea of the content of this notion [of computable functional], but not that the axioms [given in this paper] hold for it. The same apparently paradoxical situation also obtains for the notion, basic to intuitionistic logic, of a proof that is informally understood to be correct.

The point Gödel makes in this footnote is reminiscent of paragraph XXIV in Leibniz' *Discours de Métaphysique*, where attention is drawn to the fact that we may have 'clear' or even 'distinct' knowledge of things of which we do not yet have (and may well never have) a complete analyis into primitive terms.[193] Whether Gödel had this passage in Leibniz, which he surely knew, in mind when writing this footnote, remains an open question.

The philosophical position Gödel takes up in 1958 also explains why, unlike in 1933 and 1941, he no longer actually questions that it would be appropriate to consider the Proof Interpretation constructive, but only claims higher degrees of evidence and constructiveness for his interpretation. For example, Gödel's objections to the intuitionistic notion of proof from 1933 and 1941 that we quoted above would equally apply to the primitive notion he substitutes for it in 1958. By that time he had relaxed the stringent conditions he had put on constructivity in the 1930s, and, correspondingly, had come to see room for a constructive but 'hitherto unutilized extension of the finitary standpoint'.

Regarding the epistemological advantages over the Proof Interpretation that Gödel claims in 1958, an intuitionist could say the following: 'If one considers the Proof Interpretation, not in general, but limited to a particular topic, it is only to be expected that it can be turned into something more specific; the computable functionals are an example of just this for the case of arithmetic.'[194] This idea was used above when motivating the clause for the translation of implications. (Kleene's realizability interpretations of HA may be motivated in the same way, but a difference is that the Dialectica Interpretation aims for interpretations that

[191] [Wang, 1974, pp.84–85]
[192] [Gödel, 1990, p.245n.5]
[193] [Leibniz, 1880, p.449], where also the terms 'clear' and 'distinct' are defined.
[194] See also pp.231–232 of Troelstra's introduction in [Gödel, 1990].

are more evident than the original statements.) Gödel does not go into this issue, but seems to leave open the possibility: 'Obviously, we do not claim that [the clauses of the translation] reproduce the meaning of the logical particles introduced by Brouwer and Heyting. Further investigation is needed to see how far these can be replaced by our definitions.'[195]

In 1972, Gödel pointed out that the Dialectica Interpretation validates MP in the form (equivalent to $\mathrm{MP}_{\mathrm{PR}}$ as stated above)

$$\neg\forall x A(x) \to \exists x \neg A(x)$$

for primitive recursive A; he adds that, in a more general setting, the principle is provable for any decidable property A of any objects x, and he takes this as an independent argument for the higher degree of constructiveness of the Dialectica Interpretation.[196] But as Troelstra points out,[197] this argument seems doubtful because of the combined facts that MP is refutable in the theory of lawless sequences (a decidable notion)[198] and that Gödel considers such sequences as close to Hilbert's finitism.[199] The matter is quite delicate. In a draft letter to Bernays from July 1969, Gödel reports that 'it now seems to me, after more careful consideration, that choice sequences are something concretely evident and therefore are finitary in Hilbert's sense, even if Hilbert himself was perhaps of another opinion'.[200] He also plans to include a footnote to this effect in his revision of the 1958 paper.[201] In a letter to Bernays of July 25, 1969, which he actually sent, he says that choice sequences seem to him 'to be quite concrete, but not to extend finitism in an essential way'.[202]

As functionals of a given type may well be defined by referring to (arbitrarily) more complex types, the Dialectica Interpretation seems to be highly impredicative. Something similar could be suggested about the Proof Interpretation, as proofs may involve references to (arbitrarily) more complex proofs. However, this worry may be alleviated if one accepts that the notion of computable functional (respectively, intuitionistic proof) is truly primitive and that it suffices that one recognize one when one sees one. We will not go further into this issue here, and refer to the literature.[203]

[195] [Gödel, 1990, pp.249–251]
[196] [Gödel, 1990, p.276, footnote h]
[197] [Gödel, 1990, p.232]
[198] MP here is taken in the form $\neg\neg\exists x(\alpha(x) = 0) \to \exists x(\alpha(x) = 0)$; the lawless character of α ensures both that the antecedent is true and that no proof of the consequent can be extracted from this.
[199] [Gödel, 1990, p.272, footnote c]
[200] [Gödel, 2003a, p.269]
[201] [Gödel, 2003a, p.269–270]
[202] [Gödel, 2003a, p.271]. See also [Kreisel, 1980, p.202].
[203] E.g. Bernays' letter to Gödel of March 16, 1972 [Gödel, 2003a, pp.294–295], Gödel's letter to Bernays of December 26, 1972 [Gödel, 2003a, pp.298–301], [Kreisel, 1987, p.111], Troelstra's introduction to both versions of the Dialectica paper [Gödel, 1990, pp.233, 235–236], [Tait, 2001], and, in particular, [Dummett, 2000, pp.269–274].

In his lectures and writings on functional interpretations, Gödel showed, as he also did in his paper on Kant,[204] his talent for exploring and further developing a philosophical position (here, the finitary position) that is not his own.

The Dialectica Interpretation and its offspring have found many applications in proof theory.[205]

Influences of intuitionism on Gödel's work

In the draft of an unsent letter to Dirk van Dalen from 1969, Gödel wrote:[206]

> My relationship with Intuitionism consists primarily in some theorems I proved about certain parts of intuitionistic mathematics in particular that published in Dial[ectica] 12. The question as to whether this paper is important for the *foundations* of Intuitionism I must leave for Intuitionists to answer. I did not write the paper from this point of view and some supplementation would be necessary in order to clarify it's [sic] relevance for the foundations of Intuitionism. [Gödel's emphasis.]

The view on the Dialectica Interpretation expressed here seems to be valid for Gödel's work on intuitionistic systems in general. He recognized the epistemological advantages of constructivism, and looked for interpretations of formal systems for intuitionistic logic and arithmetic. What he did not do was to explore and develop the interpretations intended by the intuitionists themselves. As intuitionistic analysis involves objects (choice sequences) that are not accepted in classical mathematics (because they depend for their existence on the mathematician's mind), this may also explain why Gödel never worked or lectured in that domain.[207] This is a further illustration of the observation made above on Brouwer's proof of the bar theorem: to Gödel this way of doing mathematics must amount to making the notions of mathematical truth and existence dependent on something that they should not be. Indeed, according to Sue Toledo's notes of her conversations with Gödel in the mid-1970s, he explained that

> [in intuitionism] statements involve [an] extra-mathematical element. Namely, the mind of the mathematician + his ego. Statements of int[uitionism] are psychol[ogical] statements, but not of empirical psy[chology],—essential a priori psychology / not formal.[208]

Yet, his view that intuitionism involves extra-mathematical elements (which, as will be clear, also induce various constraints compared to classical mathematics) did not keep Gödel from getting something out of Brouwer's typically intuitionistic

[204] [Gödel, *1946]

[205] For an overview and references, see [Gödel, 1990, pp.238–239], [Feferman, 1998] and [Avigad and Feferman, 1999].

[206] Gödel Nachlaß, Firestone Library, Princeton, item 012891.

[207] In particular, the lectures on intuitionism (Princeton 1941?) mentioned on p. 537 of [Gödel, 2003b] do not discuss analysis.

[208] [Toledo, unpublished, p.42]

thinking; we mention the following three cases where Gödel's concrete work in logic was directly influenced by Brouwer's ideas.

In conversation with Hao Wang, Gödel claimed, 'In 1942 I already had the independence of the axiom of choice [in finite type theory]. Some passage in Brouwer's work, I don't remember which, was the initial stimulus.'[209] In Arbeitsheft 14, which contains his notes on the proof from 1942, one sees that Gödel used Brouwer's continuity principle for choice sequences to define a notion of 'intuitionistic truth' for propositions about infinite sequences (of natural numbers).[210] Choice sequences are potentially infinite objects and depend in their development on the mathematician's choices; as a consequence, one generally has no complete information about them (except in the case where the mathematician commits herself to following an algorithm). On this, Brouwer based his continuity principle, which states that if to every choice sequence a natural number is assigned, then for each sequence this number is already determined by an initial segment of it (which is finite).[211] By 1942, Gödel may have seen the statement of the principle in [Brouwer, 1918, p.13], [Brouwer, 1924, p.189] or [Brouwer, 1927, p.63].[212] In his proof, Gödel thinks of classical sets as given by sequences, and treats these sequences as if they were choice sequences. Gödel also said to Wang that the method he used in 1942 was 'related' to Cohen's.[213] Behind that remark must be the fact that in Cohen's forcing, too, truth values for propositions (in an appropriately restricted language, here the first-order language of pure set theory) about certain infinite objects (here, generic sets) are always already determined by information about a finite part of such an object. Forcing therefore involves a notion of continuity analogous to that for choice sequences; this analogy explains why the logic imposed by choice sequences, and the logic that corresponds to the clauses of the definition of forcing, are the same, i.e. intuitionistic logic.[214] (To point out this analogy is of course not to suggest that Gödel invented forcing be-

[209] [Wang, 1996, p.86]. Gödel's attempts in 1966 and 1967 to reconstruct this proof from his notes failed; see his letters to Alonzo Church of September 29, 1966 [Gödel, 2003a, p.372] and to Wolfgang Rautenberg of June 30, 1967 [Gödel, 2003b, p.183]. Charles Parsons [Gödel, 2003b, p.180] has drawn attention to the following. To Church, Gödel only said that he had an independence proof of the axiom of constructibility in type theory but added that 'I believed (and I think rightly) that I could extend my proof to this case [i.e., the axiom of choice] without serious difficulty'; a year later, to Rautenberg he claimed, as he would do later to Wang, that he actually had been in possession of an independence proof of the axiom of choice in type theory.

[210] Arbeitsheft 14 is item 030032 in the Gödel Nachlaß, Firestone Library, Princeton. See e.g. p.14 (clause 1), p.15 (clause 6), p.16 (in 'Df. 'W").

[211] For a discussion of problems and possibilities in justifications of this principle, see [van Atten and van Dalen, 2002].

[212] Reading notes referring to the latter two papers are in item 050066 in the Gödel Nachlaß, Firestone Library, Princeton. These notes are not dated. On the otherwise empty back Gödel wrote 'Brouwer bar theorem'; that English term was introduced only in [Brouwer, 1954]. But it is not excluded that Gödel made these notes before or in 1942 and then added that jotting on the back later.

[213] [Wang, 1996, p.251]

[214] For continuity as a motivation of forcing, see [Wang, 1981, section 7.11]. On the theme of interpretations of intuitionistic logic and forcing, see for example [Kripke, 1965, pp.118–120], [Vopěnka, 1965] and [Fitting, 1969].

fore Cohen: much more than the idea of finite approximations is needed to arrive at that.) Finally, Gödel also described to Wang 'a metamathematical consistency proof of the axiom of choice' that he had obtained in 1940.[215] Of that proof, which Gödel did not manage to reconstruct for Wang from his old notes, he said that 'it uses finite approximations. Only the initial segment is known. Ask whether, say, 2 will occur: yes or no', and this again points to Brouwer's continuity principle.

The second case of Brouwer influencing Gödel is of a heuristic rather than a technical nature and concerns the (first) incompleteness theorem. In 1928, Brouwer gave two lectures in Vienna: on March 10, on general philosophy and intuitionistic foundations of mathematics; and on March 14, on the intuitionistic theory of the continuum.[216] According to Hao Wang, 'it appears certain that Gödel must have heard the two lectures'; indeed, Gödel wrote to Menger on April 20, 1972 that 'I only saw [Wittgenstein] once in my life when he attended a lecture in Vienna. I think it was Brouwer's'.[217] An entry in Carnap's diary for December 12, 1929, states that Gödel talked to him that day 'about the inexhaustibility of mathematics (see separate sheet) He was stimulated to this idea by Brouwer's Vienna lecture. Mathematics is not completely formalizable. He appears to be right'.[218] On the 'separate sheet', Carnap wrote down what Gödel had told him:

> We admit as legitimate mathematics certain reflections on the grammar of a language that concerns the empirical. If one seeks to formalize such a mathematics, then with each formalization there are problems, which one can understand and express in ordinary language, but cannot express in the given formalized language. It follows (Brouwer) that mathematics is inexhaustible: one must always again draw afresh from the 'fountain of intuition'. There is, therefore, no characteristica universalis for the *whole* mathematics, and no decision procedure for the whole mathematics. In each and every *closed language* there are only countably many expressions. The *continuum* appears only in 'the whole of mathematics' ... If we have *only one language*, and can only make 'elucidations' about it, then these elucidations are inexhaustible, they always require some new intuition again.[219]

This record contains in particular elements from the second of Brouwer's lectures, in which one finds the argument that Gödel refers to: on the one hand, the full continuum is given in a priori intuition, while on the other hand, it cannot be exhausted by a closed language with countably many expressions.

[215] [Wang, 1996, p.86]

[216] Published as [Brouwer, 1929] and [Brouwer, 1930].

[217] [Wang, 1987, p.80] and [Gödel, 2003b, p.133]. In a letter of January 19, 1967 to George W. Corner of the American Philosophical Society (carbon copy in Gödel-Nachlass, item 021257), Gödel, requested to write a biography of Brouwer, wrote that 'I have seen Brouwer only on one occasion, in 1953, when he came to Princeton for a brief visit'; perhaps Gödel meant that he had never actually talked to Brouwer before 1953 (more on that visit below).

[218] [Wang, 1987, p.84]

[219] As quoted, in translation, in [Wang, 1987, p.50].

In fact, Brouwer must have realized already around 1907 that one can diagonalize out of formal systems. In his dissertation of that year he noted that the totality of all possible mathematical constructions is 'denumerably unfinished'; by this he meant that 'we can never construct in a well-defined way more than a denumerable subset of it, but when we have constructed such a subset, we can immediately deduce from it, following some previously defined mathematical process, new elements which are counted to the original set'.[220] In one of the notebooks leading up to his dissertation, Brouwer stated that 'The totality of mathematical theorems is, among other things, also a set which is denumerable but never finished'.[221] This explains why Brouwer could say (e.g. to Freudenthal and to Wang) that he was not surprised by Gödel's (first) incompleteness theorem.[222]

Of course, in the theorems that Gödel, after having been inspired by Brouwer's lectures, eventually arrived at, he went considerably beyond Brouwer. As Gödel stressed (without reference to Brouwer's lectures) in his letter to Zermelo of October 12, 1931:

> I would still like to remark that I see the essential point of my result not in that one can somehow go outside any formal system (that follows already according to the diagonal procedure), but that for every formal system of metamathematics there are statements which are *expressible* within the system but which may *not be decided* from the axioms of that system, and that those statements are even of a relatively simple kind, namely, belonging to the theory of the positive whole numbers.[223]

Brouwer never explicitly commented on Gödel's theorems in print, but clearly had the (second) incompleteness theorem in mind when he wrote, in 1952, that

> The hope originally fostered by the Old Formalists that mathematical science erected according to their principles would be crowned one day with a proof of non-contradictority, was never fulfilled, and, nowadays, in view of the results of certain investigations of the last few decades, has, I think, been relinquished.[224]

And when Brouwer admitted, in 1955, that 'the sumptuous symbolic logic of the twentieth century ... at present is continually raising the most captivating problems and making the most surprising and penetrating discoveries',[225] it is more than likely that he was thinking of, among other things, Gödel's work.

[220] [Brouwer, 1907, p.82]

[221] Brouwer, Notebook VIII, p.44. Brouwer Archive, Dept. of Philosophy, Utrecht; translation from the Dutch ours. By 'theorem', Brouwer will not have meant a derivation in a(n interpreted) formal system, but rather a report (by the ideal mathematician) on a successful mathematical construction in the mind.

[222] For Freudenthal, see [van Dalen, 2001, p.359]; for Wang, [Wang, 1987, p.88].

[223] [Gödel, 2003b, p.429]; original emphases.

[224] [Brouwer, 1952, p.508]. As examples of 'Old Formalists' Brouwer mentions Dedekind, Cantor, Peano, Hilbert, Russell, Zermelo, and Couturat. Note the qualification 'I think'.

[225] [Brouwer, 1955, p.554]

The third case of an intuitionistic notion that Gödel had use for is that of the weak (or 'Brouwerian') counterexample; for instance, intuitionistically, there is no reason to assume $\forall x \in \mathbb{R}(x \in \mathbb{Q} \vee x \notin \mathbb{Q})$.[226] This notion was illustrated by Brouwer in the first Vienna lecture; it inspired Gödel in his refutation in 1930 of Behmann's claim, advanced earlier that year, that classical existence proofs (not involving the uncountable infinite) can always be made constructive.[227] It is in their reconstruction of analysis that intuitionists arrive at so-called 'strong counterexamples' to principles of classical logic. For example, Brouwer showed that actually $\neg\forall x \in \mathbb{R}(x \in \mathbb{Q} \vee x \notin \mathbb{Q})$.[228] Such phenomena are possible because of the interaction between ontology and logic that occurs in intuitionistic analysis; but Gödel always stayed away from characteristically intuitionistic ontology. This last remark would need some qualification to the (apparently varying) extent that Gödel was ready to accept lawless sequences as objects of finitary mathematics (see the discussion of the Dialectica Interpretation, above). Note that to accept them as such would imply that finitary mathematics is not a proper part of classical mathematics. In so far as finitary mathematics is conceived of as a (schematic) theory of our handling of intuitive objects, this may indeed be the case.[229] We will not develop this theme further here; but we note that this would go together well with Gödel's view, as reported in a letter from Shen Yuting to Hao Wang of April 3, 1974, that 'classical mathematics does not "include" constructive mathematics'.[230]

For all the differences between platonism and intuitionism, Gödel and Brouwer also had certain philosophical views in common. Both held that mathematics has content, and in particular describes entities that are not of a linguistic nature. It is therefore not surprising that their respective views on the role of language in mathematics are quite similar.[231] Language has a practical, supporting role to play, but in no way determines the nature of mathematics itself. This holds in particular for formal systems; see for example, Gödel's reaction to Cohen's work.[232] Kreisel reports that, in October 1955, Gödel spoke to him of the 'Aussichtlosigkeit' of 'doing anything decisive in foundations by means of mathematical logic';[233] and in a note from the period May 1941–April 1942, Gödel had already remarked:

> Perhaps the reason why no progress is made in mathematics (and there are so many unsolved problems), is that one confines oneself to the ext[ensional]—thence also the feeling of disappointment in the case of many theories, e.g. propositional logic and formalisation altogether.[234]

[226] [Brouwer, 1929, section III]
[227] [Gödel, 2003a, p.17], [Gödel, 2003b, pp.565–567] and [Mancosu, 2002].
[228] [Brouwer, 1928, §2]
[229] [van Atten, 2004b, p.425]
[230] Gödel Nachlaß, Firestone Library, Princeton, item 013133; compare [Kreisel, 1980, p.202].
[231] For a statement of Brouwer's views, see e.g. [Brouwer, 1907, ch.3], [Brouwer, 1929] and the introduction to [Heyting, 1930a]; for Gödel's, e.g. [Wang, 1996, pp.180–181]
[232] [Gödel, 2003b, p.372, letter 6]; see also [Kennedy and van Atten, 2004, section 5.3].
[233] [Kreisel, 1987, p.107]; p.104 for the date.
[234] Notebook Max IV, p.198 (Gödel Nachlaß, Firestone Library, Princeton, item 030090). Transcription Cheryl Dawson; translation from the German ours; amendment ours.

On both the platonist and the intuitionist view, foundational progress will have to come mainly from informal analysis of intuitive concepts. To that end, Gödel eventually made an explicit turn to phenomenology as a method;[235] Brouwer did not, but his work easily lends itself to phenomenological interpretation.[236] (This suggests a common domain in which to describe and, perhaps, in some sense resolve the obvious conflict between their foundational views. Ultimately, the conflict between Gödel's platonism and Brouwer's intuitionism is one between two kinds of idealism, not unlike the conflict between the views of Kant and Hegel.[237]) Given this emphasis on the intuitive and the intensional, it is not surprising that, as Kreisel informed us, Gödel encouraged his developing and advocating the notion of informal rigour—at the same time warning him that mathematicians would not be enamoured of the idea.[238] Gödel was pleased by its successful application to the intuitionistic notion of lawless sequence, and hoped for a similar success when applied to the classical notion of set (here Kreisel was more skeptical).[239]

Gödel's contacts with Brouwer and Heyting

As we saw, Gödel derived significant stimulation for his technical work from some of Brouwer's lectures and papers. On a more personal level, little love seems to have been lost between the two. When Brouwer visited Princeton in 1953, Gödel invited him once for lunch and once for tea. From Gödel's remarks on these visits in a letter to his mother of October 31, 1953,[240] one gathers that Gödel only did this because he felt obliged to. Indeed, Kreisel reports that 'Gödel was utterly bored by Brouwer', in spite of the latter's 'probably genuine exuberance'.[241] Brouwer, in a letter to Morse of January 4, 1955,[242] sent his best wishes to several people at the Institute for Advanced Study, but did not include Gödel. A more positive, though less direct, connection between Brouwer and Gödel is that Hendrik Poutsma, the author of a handbook on English grammar that Gödel considered authoritative,[243] was an uncle of Brouwer's.

Gödel first met Brouwer's former student and then foremost collaborator, Arend Heyting, at the Königsberg conference, where the latter gave a talk on the intuitionistic foundations of mathematics.[244] When Heyting, soon after, received from Gödel an offprint of the incompleteness paper, he wrote back, 'Little could I imagine in Königsberg that your remark that there could be means of proof in intuitionistic mathematics that do not occur in Hilbert's system concealed such a

[235] [Gödel, *1961/?] and its introduction, [Tieszen, 1998], and [van Atten and Kennedy, 2003].
[236] See e.g. [van Atten, 2004a], especially ch.6, and [van Atten *et al.*, 2002].
[237] See [van Atten and Kennedy, 2003], in particular section 3.2.
[238] Letter of Kreisel to van Atten, January 10, 2005. For the notion of 'informal rigour', see [Kreisel, 1967]; the term, according to the letter referred to, is Kreisel's.
[239] [Kreisel, 1987, p.166] and [Troelstra, 1985].
[240] [Schimanovich-Galidescu, 2002, p.197]
[241] [Kreisel, 1987, p.146]
[242] Brouwer Archive, Dept. of Philosophy, Utrecht.
[243] [Poutsma, 1914–1929]; Gödel [2003b, p.303] refers to the second edition of vol.2.
[244] [Heyting, 1931]

deep background'.[245] Like Brouwer, Heyting will not have been surprised by the incompleteness of formal systems for arithmetic, but Heyting acknowledged more explicitly the work behind it.

Plans in the early 1930s for a joint book by Heyting and Gödel, which was to present an overview of contemporary research in the foundations of mathematics, never quite materialized. Eventually, Heyting published his part separately,[246] and Gödel never completed his.[247]

In December 1957, Gödel and Heyting met again, when Heyting came to Princeton to give two lectures at the Institute for Advanced Study and one at Princeton University; the former were on Gödel's invitation.[248] In his invitation letter, Gödel expressed the hope that Heyting would 'be able to stay in Princeton for some days in addition to those when you will be giving the lectures so that we may discuss foundational questions with you'.[249] Because of Heyting's further commitments, he actually had to leave Princeton immediately. But according to a little diary that he kept of his American tour,[250] on December 9 Gödel and he did have a discussion on impredicative definitions. This may have included a discussion of the clause for implication in the Proof Interpretation; no further notes on that conversation seem to exist.[251] Perhaps Gödel also informed Heyting of his result on Markov's Principle (see above), but, again, no confirmation has been found that he did. A third topic they could have discussed, but apparently did not (and which was not addressed in Heyting's three lectures), is the remarkably similar views on the relations between various forms of constructive or intuitionistic mathematics and classical mathematics, in terms of a scale of degrees of evidence, that both had arrived at in the almost three decades since their meeting in Königsberg; compare Gödel's lecture for the Mathematical Association of America from 1933[252] with Heyting's inaugural lecture from 1949.[253] The most important difference between Gödel and Heyting is that Gödel embraces the metaphysical component in classical mathematics (and logic) while Heyting eschews it. One can only wish that their renewed acquaintance would have led to, finally, joint work.

[245] [Gödel, 2003b, p.35]

[246] [Heyting, 1934]

[247] For the full story, see the many exchanges between Heyting and Gödel, as well as Parsons' introduction to their correspondence, in [Gödel, 2003a].

[248] The lectures were: 'Intuitionistic theory of measure and integration', IAS, December 9; 'The interpretation of intuitionistic logic', IAS, December 10; and 'On the fundamental ideas of intuitionism' in the Louis Clark Vanuxem Lectures series, Princeton University, December 11. In his lecture on intuitionistic logic, Heyting restricts himself to discussing the 'originally intended interpretation', i.e., the Proof Interpretation; in particular, he does not discuss Gödel's work. So far, these lectures have not been published, but their manuscripts are kept in the Heyting Nachlaß, Rijksarchief Noord-Holland, Haarlem, as items V57B, V57A and V57, respectively.

[249] Gödel to Heyting, October 7, 1957 (Heyting Nachlaß, item V57E-b-6).

[250] Heyting Nachlaß, item V57E-r.

[251] The diary also mentions an otherwise unspecified conversation with Gödel in the entry for December 6–7 and lunch with Gödel on December 8.

[252] [Gödel, *1933o, p.51, first whole paragraph]

[253] [Heyting, 1949, p.13]; for an English translation of the relevant passage, see [Kennedy and van Atten, 2004, p.322].

The last contact between Heyting and Gödel seems to have been in 1969, when Heyting inquired if Gödel were interested, as was rumoured, in publishing his collected works. If true, Heyting continued, he would very much like to have them appear in the series *Studies in Logic* (North-Holland), of which he was one of the editors. But Gödel replied that he actually had no such interest, and that he considered such a project not very useful, as his important papers were all readily available.[254]

ACKNOWLEDGEMENTS.

Thanks are due to Jouko Väänänen for discussions of the material in part one, particularly in clarifying some of the technical issues relevant to Gödel's argument in section 1.2.

For comments and discussion, we are grateful to Sam Buss, Wilfrid Hodges, Roman Kossak, Per Martin-Löf, Wilfried Sieg, Göran Sundholm, Robert Tragesser, Anne Troelstra, and Dirk van Dalen. Two interviews with Georg Kreisel (February 2005), as well as correspondence with him, have been helpful in further developing our own views on Gödel.

Many thanks to the staff of the Department of Rare Books and Special Collections at the Firestone Library of Princeton University, to Marcia Tucker of the Library of the Institute for Advanced Study, to the staff of the Rijksarchief Noord-Holland, Haarlem (Heyting Nachlaß), and to Dirk van Dalen and John Kuiper at the Brouwer Archive, Utrecht, for facilitating our research.

Cheryl Dawson kindly made available to us her transcriptions of Gödel's notebook Max IV and of Arbeitsheft 14.

For permission to quote unpublished material from the sources mentioned, we thank the Institute for Advanced Study (Gödel *Nachlaß*), the Brouwer Archive, Sue Toledo, and Cheryl Dawson.

BIBLIOGRAPHY

[Artemov, 2001] S. Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, 7(1):1–36, 2001.

[Avigad and Feferman, 1999] J. Avigad and S. Feferman. Gödel's functional ('Dialectica') interpretation. In S. Buss, editor, *Handbook of proof theory*, pages 337–405. North-Holland, Amsterdam, 1999.

[Badesa, 2004] C. Badesa. *The birth of model theory.* Princeton University Press, Princeton, NJ, 2004. Löwenheim's theorem in the frame of the theory of relatives, Translated from the Spanish by Michael Maudsley.

[Barwise, 1977] J. Barwise, editor. *Handbook of mathematical logic.* North-Holland Publishing Co., Amsterdam, 1977. With the cooperation of H. J. Keisler, K. Kunen, Y. N. Moschovakis and A. S. Troelstra, Studies in Logic and the Foundations of Mathematics, Vol. 90.

[Barzin and Errera, 1927] M. Barzin and A. Errera. Sur la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin de la classe des sciences*, 13:56–71, 1927.

[254] [Gödel, 2003b, pp.74–75]

[Bazhanov, 2003] V.A. Bazhanov. The scholar and the 'Wolfhound Era': The fate of Ivan E. Orlovs ideas in logic, philosophy, and science. *Science in context*, 16(4):535–550, 2003.

[Becker, 1927] O. Becker. Mathematische Existenz. Untersuchungen zur Logik und Ontologie mathematischer Phänomene. *Jahrbuch für Philosophie und phänomenologische Forschung*, VIII:439–809, 1927.

[Becker, 1930] O. Becker. Zur Logik der Modalitäten. *Jahrbuch für Philosophie und phänomenologische Forschung*, 11:497–548, 1930.

[Behmann, 1922] H. Behmann. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Mathematische Annalen*, 86:419–432, 1922.

[Benacerraf and Putnam, 1983] P. Benacerraf and H. Putnam, editors. *Philosophy of mathematics: selected readings* (2nd ed.). Cambridge University Press, Cambridge, 1983.

[Bernays, 1935] P. Bernays. Sur le platonisme en mathématique. *L'enseignement mathématique*, XXIV(1–2):52–69, 1935. Page reference to English translation 'On platonism in mathematics' in [Benacerraf and Putnam, 1983, pp.258–271].

[Bernays, 1967] P. Bernays. Hilbert, David. In P. Edwards, editor, *The encyclopedia of philosophy*, volume 3, pages 496–504. Macmillan and The Free Press, New York, 1967.

[Blass, 1998] A. Blass. Some semantical aspects of linear logic. *J. Interest Group in Pure and Applied Logic*, 5:115–126, 1998.

[Bolzano, 1992] B. Bolzano. *Bernard Bolzano—Gesamtausgabe. Reihe I. Schriften. Band 13. Teil 3.* Friedrich Frommann Verlag Günther Holzboog GmbH & Co., Stuttgart, 1992. Wissenschaftslehre. §§349–391. [Theory of science. Sections 349–391], Edited and with an introduction by Jan Berg.

[Brouwer, 1907] L.E.J. Brouwer. *Over de grondslagen der wiskunde*. PhD thesis, University of Amsterdam, 1907. Page references to English translation 'On the foundations of mathematics' [Brouwer, 1975, pp.13–101].

[Brouwer, 1918] L.E.J. Brouwer. Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten. Erster Teil, Allgemeine Mengenlehre. *KNAW Verhandelingen*, 5:1–43, 1918. Also in [Brouwer, 1975, pp.150–190].

[Brouwer, 1924] L.E.J. Brouwer. Bewijs dat iedere volle functie gelijkmatig continu is. *KNAW Verslagen*, 33:189–193, 1924. English translation 'Proof that every full function is uniformly continuous' [Mancosu, 1998, pp.36–39].

[Brouwer, 1927] L.E.J. Brouwer. Über Definitionsbereiche von Funktionen. *Mathematische Annalen*, 97:60–75, 1927. English translation [van Heijenoort, 1967, pp.457–463].

[Brouwer, 1928] L.E.J. Brouwer. Intuitionistische Betrachtungen über den Formalismus. *KNAW Proceedings*, 31:374–379, 1928. Also [Brouwer, 1975, pp.409–414]. References are to the English translation [Mancosu, 1998, pp.40–44].

[Brouwer, 1929] L.E.J. Brouwer. Mathematik, Wissenschaft und Sprache. *Monatshefte für Mathematik und Physik*, 36:153–164, 1929. Also [Brouwer, 1975, pp.417–428]. English translation [Mancosu, 1998, pp.45–53].

[Brouwer, 1930] L.E.J. Brouwer. *Die Struktur des Kontinuums*. Komitee zur Veranstaltung von Gastvorträgen ausländischer Gelehrter der exakten Wissenschaften, Wien, 1930. Also [Brouwer, 1975, pp.429–440]. English translation [Mancosu, 1998, pp.54–63].

[Brouwer, 1952] L.E.J. Brouwer. Historical background, principles and methods of intuitionism. *South African journal of science*, 49:139–146, 1952. Quoted from [Brouwer, 1975, pp.508–515].

[Brouwer, 1954] L.E.J. Brouwer. Points and spaces. *Canadian journal of mathematics*, 6:1–17, 1954. Cited from [Brouwer, 1975, pp.522–538].

[Brouwer, 1955] L.E.J. Brouwer. The effect of intuitionism on classical algebra of logic. *Proceedings of the Royal Irish Academy*, 57:113–116, 1955. Quoted from [Brouwer, 1975, pp.551–554].

[Brouwer, 1975] L.E.J. Brouwer. *Collected works I. Philosophy and foundations of mathematics* (ed. A. Heyting). North-Holland, Amsterdam, 1975.

[Buss, 1994] S.R. Buss. On Gödel's theorems on lengths of proofs. I. Number of lines and speedup for arithmetics. *J. Symbolic Logic*, 59(3):737–756, 1994.

[Church, 1928] A. Church. On the law of the excluded middle. *Bulletin of the American Mathematical Society*, 34:75–78, 1928.

[Cohen, 1963] P. Cohen. The independence of the continuum hypothesis. *Proc. Nat. Acad. Sci. U.S.A.*, 50:1143–1148, 1963.

[Craig, 1965] W. Craig. Satisfaction for n-th order languages defined in n-th order languages. *J. Symbolic Logic*, 30:13–25, 1965.

[Davis, 2004] M. Davis, editor. *The undecidable*. Dover Publications Inc., Mineola, NY, 2004. Basic papers on undecidable propositions, unsolvable problems and computable functions, Corrected reprint of the 1965 original [Raven Press, Hewlett, NY; MR0189996].

[Dawson, 1991] J.W. Dawson, Jr. The reception of Gödel's incompleteness theorems [see MR0964491 (89m:01068)]. In *Perspectives on the history of mathematical logic*, pages 84–100. Birkhäuser Boston, Boston, MA, 1991.

[Dawson, 1997] J.W. Dawson, Jr. *Logical dilemmas*. A K Peters Ltd., Wellesley, MA, 1997. The life and work of Kurt Gödel.

[de Swart, 1977a] H.C.M. de Swart. Another intuitionistic completeness proof. *Journal of Symbolic Logic*, 41(3):644–662, 1977.

[de Swart, 1977b] H.C.M. de Swart. An intuitionistically plausible interpretation of intuitionistic logic. *Journal of Symbolic Logic*, 42(4):564–578, 1977.

[Dehornoy, 2004] P. Dehornoy. Progrès récents sur l'hypothèse du continu (d'après Woodin). *Astérisque*, (294):viii, 147–172, 2004.

[Detlefsen, 1979] M. Detlefsen. On interpreting Gödel's second theorem. *J. Philos. Logic*, 8(3):297–313, 1979.

[Došen, 1992] K. Došen. The first axiomatization of relevant logic. *Journal of Philosophical Logic*, 21:339–356, 1992.

[Dreben *et al.*, 1963] B. Dreben, P. Andrews, and S. Aanderaa. False lemmas in Herbrand. *Bull. Amer. Math. Soc.*, 69:699–706, 1963.

[Dummett, 2000] M. Dummett. *Elements of intuitionism* (2nd, rev. edition). Clarendon Press, Oxford, 2000.

[Ehrenfeucht and Mycielski, 1971] A. Ehrenfeucht and J. Mycielski. Abbreviating proofs by adding new axioms. *Bull. Amer. Math. Soc.*, 77:366–367, 1971.

[Ewald, 1996] W. Ewald, editor. *From Kant to Hilbert: a source book in the foundations of mathematics. Vol. I, II*. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1996.

[Feferman, 1960/1961] S. Feferman. Arithmetization of metamathematics in a general setting. *Fund. Math.*, 49:35–92, 1960/1961.

[Feferman, 1962] S. Feferman. Transfinite recursive progressions of axiomatic theories. *J. Symbolic Logic*, 27:259–316, 1962.

[Feferman, 1998] S. Feferman. Gödel's Dialectica Interpretation and its two-way stretch. In *In the light of logic*, pages 209–225. Oxford University Press, Oxford, 1998. Slightly modified and updated version of the paper of the same title published in G. Gottlob et al., eds., *Computational logic and proof theory*, Lecture notes in computer science 713, Heidelberg: Springer 1993, pp.23–40.

[Fitting, 1969] M. Fitting. *Intuitionistic logic, model theory, and forcing*. North-Holland, Amsterdam, 1969.

[Gentzen, 1967] G. Gentzen. *Die Widerspruchsfreiheit der reinen Zahlentheorie*. Libelli, Band 185. Wissenschaftliche Buchgesellschaft, Darmstadt, 1967.

[Gentzen, 1969] G. Gentzen. *The collected papers of Gerhard Gentzen*. Edited by M. E. Szabo. Studies in Logic and the Foundations of Mathematics. North-Holland Publishing Co., Amsterdam, 1969.

[Glivenko, 1929] V. Glivenko. Sur quelques points de la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin de la classe des sciences*, 5(15):183–188, 1929.

[Gödel, 1931] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. Page references are to [Gödel, 1986, pp.144–195].

[Gödel, 1932] K. Gödel. Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien*, 69:65–66, 1932. Also in [Gödel, 1986, pp.222–225].

[Gödel, 1932f] K. Gödel. Zum intuitionistischen Aussagenkalkül. *Zentralblatt für Mathematik und ihre Grenzgebiete*, 2:321–322, 1932. Also in [Gödel, 1986, pp.246–247].

[Gödel, 1933] K. Gödel. [On Parry's axioms]. *Ergebnisse eines mathematischen Kolloquiums*, 4:6, 1933. Also in [Gödel, 1986, pp.266–267].

[Gödel, 1933e] K. Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums*, 4:34–38, 1933. Page references are to [Gödel, 1986, pp.286–295].

[Gödel, 1933f] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, 4:39–40, 1933. Also in [Gödel, 1986, pp.300–303].

[Gödel, 1933n] K. Gödel. Zum intuitionistischen Aussagenkalkül. *Ergebnisse eines mathematischen Kolloquiums*, 4:40, 1933. Slightly modified reprint of [Gödel, 1932].
[Gödel, *1933o] K. Gödel. The present situation in the foundations of mathematics. Lecture in Cambridge, MA (published in [Gödel, 1995, pp.45–53]), 1933.
[Gödel, *1938a] K. Gödel. Vortrag bei Zilsel. Lecture in Vienna (published in [Gödel, 1995, pp.86–113]), 1938.
[Gödel, *1941] K. Gödel. In what sense is intuitionistic logic constructive? Lecture at Yale (published in [Gödel, 1995, pp.189–200]), 1941.
[Gödel, *1946] K. Gödel. Some observations about the relationship between theory of relativity and Kantian philosophy. Manuscript (published in [Gödel, 1995, pp.230–259]), 1946.
[Gödel, 1958] K. Gödel. Über eine bisher noch nicht benutzte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287, 1958. References are to the reprint [Gödel, 1990, pp.241–251].
[Gödel, *1961/?] K. Gödel. The modern development of the foundations of mathematics in the light of philosophy. Manuscript (published in [Gödel, 1995, pp.374–387]), 1961.
[Gödel, 1972] K. Gödel. On an extension of finitary mathematics which has not yet been used. Translation and revision of [Gödel, 1958], meant for publication in *Dialectica*, eventually published in [Gödel, 1990, pp.271–280], 1972.
[Gödel, 1986] K. Gödel. *Collected Works. I: Publications 1929–1936.* (eds. S. Feferman et al.). Oxford University Press, Oxford, 1986.
[Gödel, 1990] K. Gödel. *Collected Works. II: Publications 1938–1974.* (eds. S. Feferman et al.). Oxford University Press, Oxford, 1990.
[Gödel, 1995] K. Gödel. *Collected Works. III: Unpublished essays and lectures.* (eds. S. Feferman et al.). Oxford University Press, Oxford, 1995.
[Gödel, 2003a] K. Gödel. *Collected Works. IV: Correspondence A-G.* (eds. S. Feferman et al.). Oxford University Press, Oxford, 2003.
[Gödel, 2003b] K. Gödel. *Collected Works. V: Correspondence H-Z.* (eds. S. Feferman et al.). Oxford University Press, Oxford, 2003.
[Herbrand, 1931] J. Herbrand. Sur la non-contradiction de l'arithmétique. *Journal für die reine und angewandte Mathematik*, 166:1–8, 1931. Also [Herbrand, 1968, pp.221–232]. English translation [van Heijenoort, 1967, pp.618–628].
[Herbrand, 1968] J. Herbrand. *Écrits logiques* (ed. J. van Heijenoort). Presses Unversitaires de France, Paris, 1968.
[Herbrand, 1971] J. Herbrand. *Logical writings* (ed. W. Goldfarb). Harvard University Press, Cambridge, MA, 1971.
[Hesseling, 2003] D. Hesseling. *Gnomes in the fog.* Birkhäuser, Basel, 2003.
[Heyting, 1930a] A. Heyting. Die formalen Regeln der intuitionistischen Logik I. *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, pages 42–56, 1930. English translation [Mancosu, 1998, pp.311–327].
[Heyting, 1930b] A. Heyting. Die formalen Regeln der intuitionistischen Logik II. *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, pages 57–71, 1930.
[Heyting, 1931] A. Heyting. Die intuitionistische Grundlegung der Mathematik. *Erkenntnis*, 2:106–115, 1931. Quoted from English translation [Benacerraf and Putnam, 1983, pp.52–61].
[Heyting, 1934] A. Heyting. *Mathematische Grundlagenforschung, Intuitionismus, Beweistheorie.* Springer, Berlin, 1934.
[Heyting, 1949] A. Heyting. *Spanningen in de wiskunde.* Noordhoff, Groningen, 1949. Inaugural lecture, University of Amsterdam.
[Heyting, 1956] A. Heyting. *Intuitionism, an introduction.* North-Holland, Amsterdam, 1956.
[Hilbert and Ackermann, 1972] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik.* Springer-Verlag, Berlin, 1972. Sechste Auflage, Die Grundlehren der mathematischen Wissenschaften, Band 27.
[Hilbert, 1922] D. Hilbert. Neubegründung der Mathematik: Erste Mitteilung. *Abhandlungen aus dem Seminar der Hamburgischen Universität*, 1:157–177, 1922.
[Hilbert, 1925] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95:161–190, 1925.
[Hilbert, 1965] D. Hilbert. *Gesammelte Abhandlungen. Dritter Band. Analysis, Grundlagen der Mathematik, Physik, Verschiedenes, nebst einer Lebensgeschichte.* Chelsea Publishing Co., New York, 1965.
[Kaye, 1991] R. Kaye. *Models of Peano arithmetic*, volume 15 of *Oxford Logic Guides.* The Clarendon Press Oxford University Press, New York, 1991. Oxford Science Publications.

[Kennedy and van Atten, 2004] J. Kennedy and M. van Atten. Gödel's modernism: on set-theoretic incompleteness. *Graduate Faculty Philosophy Journal*, 25(2):289–349, 2004. Erratum in 26(1).

[Kolmogorov, 1925] A.N. Kolmogorov. O principe tertium non datur. *Matematicheskii Sbornik*, 32:646–667, 1925. Quoted from English translation 'On the principle of excluded middle' [van Heijenoort, 1967, pp.416–437].

[Kolmogorov, 1932] A.N. Kolmogorov. Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift*, 35:58–65, 1932. Quoted from English translation [Mancosu, 1998, pp.328–334].

[Kreisel, 1958] G. Kreisel. Non-derivability of $\neg(x)A(x) \rightarrow (\exists x)\neg A(x)$, A primitive recursive, in intuitionistic formal systems (abstract). *Journal of Symbolic Logic*, 23(4):456–457, 1958.

[Kreisel, 1962] G. Kreisel. On weak completeness of intuitionistic predicate logic. *Journal of Symbolic Logic*, 27(2):139–158, 1962.

[Kreisel, 1967] G. Kreisel. Informal rigour and completeness proofs. In I. Lakatos, editor, *Problems in the philosophy of mathematics*, pages 138–186. North-Holland, Amsterdam, 1967.

[Kreisel, 1971] G. Kreisel. Book review. *The Collected Papers of Gerhard Gentzen*. Ed. M.E. Szabo. *Journal of Philosophy*, 68(8):238–265, 1971.

[Kreisel, 1980] G. Kreisel. Kurt Gödel. 28 April 1906–14 January 1978. *Biographical Memoirs of Fellows of the Royal Society*, pages 149–224, 1980.

[Kreisel, 1987] G. Kreisel. Gödel's excursions into intuitionistic logic. In P. Weingartner and L. Schmetterer, editors, *Gödel remembered*, pages 67–179. Bibliopolis, Napoli, 1987.

[Kripke, 1965] S. Kripke. Semantical analysis of intuitionistic logic I. In M.A.E. Dummett and J.N. Crossley, editors, *Formal systems and recursive functions*, pages pp.92–130. North-Holland, Amsterdam, 1965.

[Leibniz, 1880] G.W. Leibniz. [Discours de métaphysique]. In C.I. Gerhardt, editor, *Die philosophischen Schriften von Gottfried Wilhelm Leibniz* vol. IV, pages 427–463. Weidmann, Berlin, 1880. English translation in G.W. Leibniz, *Philosophical papers and letters* (ed. L.E. Loemker), 2nd ed., D. Reidel, Dordrecht, 1969, pp.303–330.

[Lewis and Langford, 1932] C.I. Lewis and C.H. Langford. *Symbolic logic*. Century, New York, 1932.

[Lindström, 1969] P. Lindström. On extensions of elementary logic. *Theoria*, 35:1–11, 1969.

[Lopez-Escobar and Veldman, 1975] E. Lopez-Escobar and W. Veldman. Intuitionistic completeness of a restricted second-order logic. In A. Dold and B. Eckmann, editors, *Proof theory symposion. Kiel 1974*, pages 198–232. Springer, Berlin, 1975.

[Löwenheim, 1915] L. Löwenheim. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, 76:447–470, 1915.

[Mal′cev, 1971] A. I. Mal′cev. *The metamathematics of algebraic systems. Collected papers: 1936–1967*. North-Holland Publishing Co., Amsterdam, 1971. Translated, edited, and provided with supplementary notes by Benjamin Franklin Wells, III, Studies in Logic and the Foundations of Mathematics, Vol. 66.

[Mancosu, 1998] P. Mancosu. *From Brouwer to Hilbert. The debate on the foundations of mathematics in the 1920s*. Oxford University Press, Oxford, 1998.

[Mancosu, 1999] P. Mancosu. Between Vienna and Berlin: the immediate reception of Gödel's incompleteness theorems. *History and Philosophy of Logic*, 20:33–45, 1999.

[Mancosu, 2002] P. Mancosu. On the constructivity of proofs. A debate among Behmann, Bernays, Gödel and Kaufmann. In W. Sieg, R. Sommer, and C. Talcott, editors, *Reflections on the foundations of mathematics: Essays in honor of Solomon Feferman*, pages 349–371. Association for Symbolic Logic, Urbana, 2002.

[Mancosu, 2004] P. Mancosu. Kurt Gödel, Collected Works, volumes IV and V. *Notre Dame Journal of Formal Logic*, 45:109–125, 2004.

[McCarty, 1994] D.C. McCarty. On Theorems of Gödel and Kreisel: completeness and Markov's Principle. *Notre Dame Journal of Formal Logic*, 35(1):99–107, 1994.

[McKinsey and Tarski, 1948] J.C.C. McKinsey and A. Tarski. Some theorems about the sentential calculi of Lewis and Heyting. *Journal of Symbolic Logic*, 13(1):1–15, 1948.

[Mostowski, 1966] A. Mostowski. *Thirty years of foundational studies. Lectures on the development of mathematical logic and the study of the foundations of mathematics in 1930–1964*. Acta Philosophica Fennica, Fasc. XVII. Barnes & Noble, Inc., New York, 1966.

[Mostowski, 1982] A. Mostowski. *Sentences undecidable in formalized arithmetic*. Greenwood Press, Westport, CT, 1982 (first edition 1952).

[Parikh, 1971] R. Parikh. Existence and feasibility in arithmetic. *J. Symbolic Logic*, 36:494–508, 1971.

[Peckhaus, 2002] V. Peckhaus. Oskar Beckers Stellung in der Geschichte der Modallogik. In V. Peckhaus, ed., *Die Philosophie und die Wissenschaften. Zum Werk Oskar Beckers*, pages 161–184. Wilhelm Fink Verlag, München, 2002.

[Poutsma, 1914–1929] H. Poutsma. *A grammar of late modern English, for the use of continental, especially Dutch, students* (two parts, in five volumes; two editions). Noordhoff, Groningen, 1914–1929.

[Rogers, 1967] H. Rogers, Jr. *Theory of recursive functions and effective computability*. McGraw-Hill Book Co., New York, 1967.

[Rosser, 1936] J.B. Rosser. Extensions of some theorems of Gödel and Church. *Journal of Symbolic Logic*, 1(3):87–91, 1936.

[Schimanovich-Galidescu, 2002] M.E. Schimanovich-Galidescu. Archivmaterial zu Gödel's Wiener Zeit, 1924–1940. In *Kurt Gödel. Wahrheit und Beweisbarkeit. Band 1: Dokumente und historische Analysen*, pages 135–147. öbv & hpt, Wien, 2002.

[Sieg, 1988] W. Sieg. Hilbert's Program sixty years later. *J. Symbolic Logic*, 53(2):338–348, 1988.

[Sieg, 1990] W. Sieg. Relative consistency and accessible domains. *Synthese*, 84(2):259–297, 1990.

[Sieg, 1999] W. Sieg. Hilbert's programs: 1917–1922. *Bull. Symbolic Logic*, 5(1):1–44, 1999.

[Simpson, 1985] S. G. Simpson. Reverse mathematics. In *Recursion theory (Ithaca, N.Y., 1982)*, volume 42 of *Proc. Sympos. Pure Math.*, pages 461–471. Amer. Math. Soc., Providence, RI, 1985.

[Skolem, 1920] T. Skolem. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Skrifter utgit av Videnskappsselskapet i Kristiania*, 1920.

[Skolem, 1923] T. Skolem. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. *Matematikerkongressen i Helsingfors 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, 1923.

[Skolem, 1928] T. Skolem. Über die mathematische Logik. *Norsk Matematisk Tidskrift*, 10:125–142, 1928.

[Skolem, 1970] T. Skolem. *Selected works in logic*. Edited by Jens Erik Fenstad. Universitetsforlaget, Oslo, 1970.

[Statman, 1981] R. Statman. Speed-up by theories with infinite models. *Proc. Amer. Math. Soc.*, 81(3):465–469, 1981.

[Sundholm and van Atten, in print] G. Sundholm and M. van Atten. The proper explanation of intuitionistic logic: on Brouwer's demonstration of the Bar Theorem. In M. van Atten et al., eds. *One hundred years of Intuitionism (1907–2007)*, pages 60–77. Birkhäuser, Basel, in print.

[Tait, 2001] W.W. Tait. Gödels unpublished papers on foundations of mathematics. *Philosophia Mathematica*, 9:87–126, 2001.

[Tait, 2005] W.W. Tait. *The provenance of pure reason*. Logic and Computation in Philosophy. Oxford University Press, New York, 2005. Essays in the philosophy of mathematics and its history.

[Tieszen, 1998] R. Tieszen. Kurt Gödel's path from the incompleteness theorems (1931) to phenomenology (1961). *Bulletin of Symbolic Logic*, 4(2):181–203, 1998.

[Toledo, unpublished] S. Toledo. [Notes on conversations with Gödel, 1972–1975], unpublished.

[Troelstra and van Dalen, 1988] A.S. Troelstra and D. van Dalen. *Constructivism in Mathematics* (2 vols.). North–Holland, Amsterdam, 1988.

[Troelstra, 1977] A.S. Troelstra. *Choice sequences. A chapter of intuitionistic mathematics*. Oxford University Press, Oxford, 1977.

[Troelstra, 1985] A.S. Troelstra. Choice sequences and informal rigour. *Synthese*, 62:217–227, 1985.

[Turing, 1939] A.M. Turing. Systems of logic based on ordinals. *Proceedings of the LMS, series 2*, 45:161–228, 1939.

[van Atten, 2004a] M. van Atten. *On Brouwer*. Wadsworth, Belmont, CA, 2004.

[van Atten, 2004b] M. van Atten. Review of D.E. Hesseling, 'Gnomes in the fog. The reception of Brouwer's intuitionism in the 1920s'. *Bulletin of Symbolic Logic*, 10(3):423–427, 2004.

[van Atten, 2005] M. van Atten. On Gödel's awareness of Skolem's Helsinki lecture. *History and Philosophy of Logic*, 2005.

[van Atten and Kennedy, 2003] M. van Atten and J. Kennedy. On the philosophical development of Kurt Gödel. *Bulletin of Symbolic Logic*, 9(4):425–476, 2003.

[van Atten and van Dalen, 2002] M. van Atten and D. van Dalen. Arguments for the continuity principle. *Bulletin of Symbolic Logic*, 8(3):329–347, 2002.

[van Atten *et al.*, 2002] M. van Atten, D. van Dalen, and R. Tieszen. Brouwer and Weyl: the phenomenology and mathematics of the intuitive continuum. *Philosophia Mathematica*, 10(3):203–226, 2002.

[van Dalen, 1974] D. van Dalen. Review of Herbrand: Logical writings (ed. W. Goldfarb). *Journal of Philosophy*, 71(15):544–549, 1974.

[van Dalen, 2001] D. van Dalen. *L.E.J. Brouwer 1881–1966. Een biografie. Het heldere licht van de wiskunde*. Bert Bakker, Amsterdam, 2001.

[van Heijenoort, 1967] J. van Heijenoort, editor. *From Frege to Gödel: A sourcebook in mathematical logic, 1879–1931*. Harvard University Press, Cambridge, MA, 1967.

[Veldman, 1976] W. Veldman. An intuitionistic completeness theorem for intuitionistic predicate logic. *Journal of Symbolic Logic*, 41(1):159–166, 1976.

[Vopěnka, 1965] P. Vopěnka. The limits of sheaves and applications on constructions of models. *Bulletin de l'Académie Polonaise des Sciences, série des Sciences Mathématiques, Astronomiques et Physiques*, 13:189–192, 1965.

[Wang, 1957] H. Wang. The axiomatization of arithmetic. *J. Symb. Logic*, 22:145–158, 1957.

[Wang, 1974] H. Wang. *From mathematics to philosophy*. Routledge and Kegan Paul, London, 1974.

[Wang, 1981] H. Wang. *Popular lectures on mathematical logic*. Van Nostrand Reinhold, New York, 1981.

[Wang, 1987] H. Wang. *Reflections on Kurt Gödel*. MIT Press, Cambridge, MA, 1987.

[Wang, 1996] H. Wang. *A logical journey. From Gödel to philosophy*. MIT Press, Cambridge, MA, 1996.

[Wilkie and Paris, 1987] A.J.. Wilkie and J.B. Paris. On the scheme of induction for bounded arithmetic formulas. *Ann. Pure Appl. Logic*, 35(3):261–302, 1987.

[Yourgrau, 2005] P. Yourgrau. *A world without time. The forgotten legacy of Gödel and Einstein*. Basic Books, New York, 2005.

[Zach, Fall 2003] R. Zach. Hilbert's program. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2003.

TARSKI'S LOGIC

Keith Simmons

1 INTRODUCTION

Alfred Tarski was born in Warsaw on January 14th 1902, under the family name of Tajtelbaum.[1] He studied at the University of Warsaw from 1918 to 1924, when he received his PhD in mathematics. His doctoral dissertation *O wyrazie pierwotnym logistyki* (*On the primitive term of logistic*) was written under the supervision of S. Leśniewski.[2] Appointed Docent in 1926 and later Adjunct Professor, Tarski taught at the University of Warsaw until 1939. During that year, Tarski visited the US for a lecture tour, but was prevented from returning to Poland by the outbreak of war. Between 1939 and 1942 Tarski was at Harvard University, the City College of New York, and the Institute for Advanced Study in Princeton. In 1942, he was appointed lecturer in mathematics at the University of California at Berkeley, where he remained for the rest of his career. He became professor of mathematics there in 1946, when his family joined him from Poland. In 1958, Tarski founded the Group in Logic and the Methodology of Science. He retired in 1968, though he taught for several more years, and continued his research throughout his retirement. Tarski died in Berkeley on October 27th 1983.

Tarski's work has had an enormous influence on the development of logic and mathematics over the last eighty years. He broke new ground with his work on metamathematics and semantics. His methods and results in those fields and many others — including algebra, geometry, and set theory — have become part of the fabric of modern logic and mathematics. The very divisions between these various fields appear somewhat artificial in the light of Tarski's work.

I have attempted to make this chapter accessible to the non-mathematician who is familiar with basic logic. I have focussed more on Tarski's work in logic, semantics and metamathematics, and less on Tarski's more purely mathematical work. Nevertheless, I hope that the reader will come away with a sense of Tarski's achievements in each of the many areas to which he contributed. My aim has been to make this chapter as self-contained as possible. For the reader who wants to follow out the formal details of a definition or a proof, I have endeavored to provide

[1] The family name Tarski was adopted in 1924. There is disagreement in the literature about Tarski's birthdate. I have followed Mostowski [Tarski, 1967] and the *Encyclopedia Britannica*, but other sources given the year of his birth as 1901 (see, for example, [Hodges, 1986]).

[2] The dissertation was submitted in 1923, and the essential part of it was published as the paper [Tarski, 1923]. An English translation of the paper appears in [Tarski, 1983a].

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

explanations in footnotes. And where the formal material goes beyond the scope of this chapter, I have provided further references. I quote often from Tarski, since it is very hard to improve upon Tarski's own words — no-one who is familiar with Tarski's work can fail to be struck by the clarity of his writing. Works by Tarski and others are referred to by the year of publication.[3] Throughout I have tried to provide some historical context for Tarski's work and its significance.

The chapter is divided into the following sections: (2) Metamathematics; (3) Formal definability; (4) Semantics: definability, truth, consequence, model theory; (5) Decidability and undecidability; (6) More on logic and algebra; (7) More on geometry; (8) Set theory; (9) Concluding remarks.

2 METAMATHEMATICS

We turn first to what Tarski initially called the *methodology of the deductive sciences* — though he later preferred the label "metalogic and metamathematics",[4] and often just used the term "metamathematics". Its objects of study are theories of a certain kind — *formalized deductive theories*. They constitute the subject matter of metamathematics just as

> "spatial entities constitute the subject matter of geometry and animals that of zoology" [Tarski, 1930d, p. 60].[5]

One task of metamathematics is to construct its objects of study. The construction of formalized deductive theories proceeds in accordance with the *deductive method*. When we study or advance any science, Tarski says,

> "a method would be ideal, if it permitted us to explain the meaning of every expression occurring in this science and to justify each of its assertions" [Tarski, 1941a, p. 117]

Tarski points out that this ideal can never be realized — since we must use expressions to explain the meaning of an expression, we have to take some terms as primitive, on pain of an infinite regress.

> "When we set out to construct a given discipline, we distinguish, first of all, a certain small group of expressions of this discipline that seems to us to be immediately understandable; the expressions of this group we call *primitive terms* or *undefined terms*." (p. 118)

[3] To reduce the number of footnotes, I refer to Tarski's work parenthetically in the main text. When the work of Tarski's from which I am quoting is clear from the context, I provide only a page number. For a full bibliography of Tarski, including all his papers, abstracts, monographs, exercises and problems, contributions to discussions, reviews, publications as editor, project reports and letters, see [Givant, 1986].

[4] See [Tarski, 1941a, p. 140].

[5] Compare this remark in [Tarski, 1941a]: "The methodology of the deductive sciences became a general science of deductive sciences in an analogous sense as arithmetic is the science of numbers and geometry is the science of geometrical configurations." (p. 138)

All other expressions are defined via the primitive terms. And we proceed similarly with the asserted statements of the theory:

> "Some of these statements which to us have the appearance of evidence are chosen as the so-called *primitive statements* or *axioms*..." (p. 118)

And further:

> "we agree to accept any other statement as true only if we have succeeded in establishing its validity, and to use, while doing so, nothing but axioms, definitions and such statements of the discipline the validity of which has been established previously." (p. 118)

The method of constructing a theory in accordance with these principles is the *deductive method*,[6] and the theory so constructed is a *formalized deductive theory*.[7]

So one task of metamathematics is the construction of formalized deductive theories via the deductive method. Now, as a result of the application of this method "deductive theories acquire certain interesting and important features" [Tarski, 1941a, p. 120]. It is a further task of metamathematics to uncover these features.

This further task can be carried out at two distinct levels. One can investigate specific deductive theories. In the 1920s, Tarski carried out metamathematical investigations into a number of deductive theories - for example, the sentential calculus (see [Tarski, 1930c]), the algebra of logic, the arithmetic of real numbers, the geometry of straight lines, the theory of order, and the theory of groups.[8] Tarski writes in [Tarski, 1930d]:

---

[6]Tarski points out that logic occupies a special position in regard to the deductive method. In using the method to construct any other discipline or theory, we cannot but presuppose logic. In Tarski's terminology, given a target discipline, logic will be a discipline *preceding the given discipline*. And it may be convenient to presuppose other disciplines too. Tarski writes:

> "Thus logic itself does not presuppose any preceding discipline; in the construction of arithmetic as a special mathematical discipline logic is presupposed as the only preceding discipline; on the other hand, in the case of geometry it is expedient — though not unavoidable — to presuppose not only logic but also arithmetic." [Tarski, 1941a, p. 119]

So, whenever we apply the deductive method, we must be sure to enumerate the preceding disciplines.

[7]The phrase 'formalized deductive discipline' appears in the early papers [Tarski, 1930b] and [Tarski, 1930d]. In [Tarski, 1930d], Tarski writes:

> "Naturally not all deductive systems are presented in a form suitable for objects of scientific investigation. Those, for example, are not suitable which do not rest on a definite logical basis, have no precise rules of inference, and the theorems of which are formulated in the usually ambiguous and inexact terms of colloquial language — in a word those which are not formalized." (p. 60)

[8]See [Tarski, 1983a, p. 205, fn. 2], where Tarski mentions that he investigated these last five theories during 1926–8.

> "Strictly speaking metamathematics is not to be regarded as a single theory. For the purpose of investigating each deductive discipline a special metadiscipline should be constructed." (p. 60)

But one can also proceed at a more abstract level, and explore concepts that are common to *all* these special metadisciplines. In [30d], Tarski proceeds at this higher level of generality, and we will start with Tarski's study of general deductive systems.

2.1 General deductive systems

In his paper "Fundamental Concepts of the Methodology of the Deductive Sciences" [Tarski, 1930d], Tarski's aim is

> "*to make precise the meaning of a series of important metamathematical concepts* which are common to the special metadisciplines, *and to establish the fundamental properties of these concepts.*" (p. 60, original emphases)

Tarski starts out with the fundamental concept of *consequence*, and provides the first rigorous axiomatic characterization of this notion. Tarski goes on to define a number of further concepts, including *logical equivalence*, *axiomatizability*, *independence*, *consistency*, and *completeness*. Tarski thus provided the first precise, systematic treatment of these basic metamathematical concepts.

The axiomatic system (or theory) that Tarski presents in [Tarski, 1930d] contains just two primitive concepts, *sentence* and *consequence*. Sentences are certain inscriptions of a well-defined form. And a deductive discipline is a set of (meaningful) sentences. Let S be the set of all meaningful sentences of a given language, and let A be an arbitrary set of sentences from S that compose a particular deductive discipline. '$Cn(A)$' denotes the set of consequences of set A, those sentences derived from A via rules of inference. The *schema* of a definition of consequence can be given as follows: the set of all consequences of the set A is the intersection of all sets which contain the set A and are closed under the given rules of inference. But we are working at too high a level of generality to give an exact definition of consequence; it is the task of a specific metadiscipline to establish rules of inference, and thereby an exact definition of consequence. Instead, we are to regard the notion of consequence as primitive, and characterized only through the axioms. From just these two primitive concepts, sentence and consequence,

> "almost all basic concepts of metamathematics can be defined; on the basis of the given axiom system various fundamental properties of these concepts can be established." (p. 69)

Tarski presents an axiom system composed of four axioms which express basic properties of the primitive concepts and "are satisfied in all known formalized disciplines" (p. 63). The axioms are as follows

Axiom 1. $\bar{\bar{S}} \leq \aleph_0$ (where "$\bar{\bar{S}}$" is the cardinality of the set S)

Axiom 2. If $A \subseteq S$, then $A \subseteq Cn(A) \subseteq S$.

Axiom 3. If $A \subseteq S$, then $Cn(Cn(A)) = Cn(A)$.

Axiom 4. If $A \subseteq S$, then $Cn(A) = \cup\{Cn(X) : X \text{ is a finite subset of } A\}$.

Axiom 1, Tarski says, "scarcely requires comment" (p. 63), though he does acknowledge the challenge that S cannot be denumerably infinite if sentences are regarded as physical inscriptions.[9] Axiom 2 tells us that consequences of A are sentences, and Axiom 3 tells us that consequences of consequences of A are consequences of A. According to Axiom 4, any consequence of A is a consequence of a finite subset of A — and this respects the idea that "in concrete disciplines" (p. 64) rules of inference operate on a finite number of sentences.

Tarski goes on to single out "an especially important category of sets of sentences ... namely, the *deductive systems*." A deductive system, or "*closed system*" or "simply a '*system*'" (p. 70), is a set of sentences that is closed under consequence — that is, it contains all of its consequences. Tarski writes:

> "Deductive systems are, so to speak, organic units which form the subject matter of metamathematical investigations. Various important notions, like consistency, completeness, and axiomatizability, which we shall encounter in the sequel, are theoretically applicable to any sets of sentences, but in practice are applied chiefly to systems." (p. 70)

The bulk of [Tarski, 1930d] is a detailed examination of the properties of (deductive) systems. Tarski defines a number of fundamental metamathematical concepts. Two sets of sentences are *logically equivalent* if they have all their consequences in common (p. 72). An *axiom system* of a set of sentences is a finite set which is equivalent to that set, and a set of sentences which has at least one axiom system is called *axiomatizable* (p. 72). A set of sentences is *independent* if it is not equivalent to any of its proper subsets (p. 83). A *basis of the set A* is an independent set of sentences which is equivalent to the set A (p. 88). A set of sentences is *consistent* if it is not equivalent to the set of all meaningful sentences (p. 90).[10] The *decision domain* of a set A of sentences is the set of all

[9] Of the assumption that S is denumerably infinite, Tarski writes:

> "... it may be noted here that I personally regard such an assumption as quite sensible, and that it appears to me even to be useful from a metamathematical standpoint to replace the inequality sign by the equality sign in Ax. 1." [Tarski, 1930d, p. 64]

Throughout his career, Tarski never showed any finitistic qualms.

[10] Tarski points out that his definition of consistency departs from the usual one, according to which a set of sentences is consistent if there is no sentence which together with its negation belongs to the consequences of this set. Tarski notes that the two definitions are equivalent for all disciplines based on the system of sentential calculus — indeed, this is the content of Theorem 9* of Tarski's paper [Tarski, 1930b]. Tarski also notes that his definition is much more general, since it applies to theories without negation, and to theories in which negation does not have the usual properties. Tarski cites Post as the source of this definition [Tarski, 1921].

sentences that are either consequences of A or which, when added to A, yield an inconsistent set of sentences; and A is said to be *complete* if its decision domain contains all meaningful sentences (p. 93).[11] The *cardinal degree of completeness* of the set A of sentences is the number of all systems which include the set A (pp. 100–101). And the *ordinal degree of completeness* is the smallest ordinal number π for which there is no strictly increasing sequence of type π of consistent systems which include the set A (p. 101).[12]

[Tarski, 1930d] contains proofs of a large number of seminal theorems on axiomatizability, independence, and consistency and completeness. For example, Corollary 37 (p. 87) tells us that if a system includes an infinite independent set of sentences, then it has uncountably many subsystems, of which only countably many are finitely axiomatizable. Tarski writes:

> "It is to be noted that within almost all deductive disciplines, and in particular within the simplest of them — the sentential calculus — it has been found possible to construct a set of sentences which is both infinite and independent, and thus to realize the hypothesis of the last corollary [Corollary 37]. Hence it turns out that in all these disciplines there are more unaxiomatizable than axiomatizable systems; the deductive systems are, so to speak, as a rule unaxiomatizable, although in practice we deal almost exclusively with axiomatizable systems. This paradoxical circumstance was first noticed by Lindenbaum in application to the sentential calculus." (p. 88)

(Tarski's use of the term 'paradoxical' indicates that unaxiomatizable theories came as a surprise.) For another example, Tarski establishes (by Corollary 44) that "every axiomatizable set of sentences possesses at least one basis" (p. 90). This result does not extend to unaxiomatizable sets of sentences on the basis of Axioms 1–4 (though as we will see below, it does extend to deductive disciplines that presuppose the sentential calculus). Another result, Theorem 56 (p. 98), tells us that "every consistent set of sentences can be extended to a consistent and complete system" (*ibid.*), a generalization of Lindenbaum's lemma for the sentential calculus.

2.2 Classical deductive systems

While in [Tarski, 1930d] Tarski investigated the broadest category of deductive systems, in his paper "On some fundamental concepts of metamathematics" [Tarski, 1930b], Tarski investigated a certain subcategory — the classical systems. In

[11] As with the definition of consistency, this definition of completeness departs from the usual one, and does not rest on the concept of negation. Tarski again cites Post [1921] as the source of this definition of completeness. Theorem 10* of Tarski's [1930b] states the equivalence of the two definitions for systems based on the sentential calculus.

[12] That is, there is no increasing sequence of type π of the form $\langle X, \ldots, X_\nu, \ldots, X_\eta, \ldots, \rangle$, where $X_\nu \subseteq X_\eta \subseteq S$ and $Cn(X_\nu) \neq Cn(X_\eta)$ for $\nu < \eta < \pi$. Or, in the now-standard terminology, there exists no chain of consistent theories of type π beginning with X.

[Tarski, 1930b], Tarski lays out two groups of axioms. The first contains Axioms 1–4, together with a fifth:

Axiom 5. There exists a sentence $x \in S$ such that $Cn(\{x\}) = S$.

The axioms in the second group are "of a more special nature":

> "In contrast to the first group of axioms those of the second group apply, not to all deductive disciplines, but only to those which presuppose the sentential calculus, in the sense that in considerations relating to these disciplines we may use as premises all true sentences of the sentential calculus." (p. 31)

These additional axioms contain two new primitive concepts: the conditional and negation. *Axiom 6* tells us that if x and y are meaningful sentences, so are $x \to y$ and $\neg x$. The remaining four axioms are as follows (where 'Ø' below denotes the empty set):

Axiom 7. If $X \subseteq S, y \in S, z \in S$ and $y \to z \in Cn(X)$, then $z \in Cn(X \cup \{y\})$.

Axiom 8. If $X \subseteq S, y \in S, z \in S$ and $z \in Cn(X \cup \{y\})$, then $y \to z \in Cn(X)$.

Axiom 9. If $x \in S$, then $Cn(\{x, \neg x\}) = S$.

Axiom 10. If $x \in S$, then $Cn(\{x\}) \cap Cn(\{\neg x\}) = Cn(\varnothing)$.

Axiom 7 is the rule of detachment, or modus ponens. Tarski takes Axiom 8 to be one formulation of the so-called deduction theorem, a discovery of Tarski's. Tarski writes:

> "This theorem, in its application to the formalism of *Principia Mathematica*, was first established by the author as far back as 1921 .... Subsequently the deduction theorem was often applied in metamathematical discussion."[13]

Axiom 9 is the classical law of non-contradiction, in the form "Everything follows from a contradiction". Axiom 10 is a formulation of the classical law of excluded middle.

So the axiomatic system of [Tarski, 1930b] applies to a *subcategory* of deductive systems, the *classical* deductive systems, encompassing the sentential calculus and all classical systems that extend it, including the first-order predicate calculus.

[13] Tarski [1983a, p. 32, fn. †]. Tarski reports that he established the theorem in connection with a discussion in the monograph of Ajdukiewicz [Tarski, 1921], and discussed the result in a lecture to the Warsaw Philosophical Institute, listed by title in *Ruch Filozoficzny*, vol. 6 (1921–2). As to applications of the deduction theorem, Tarski mentions that it was essential to the proofs of theorems in Lindenbaum and Tarski's note [Tarski, 1927b] and Tarski's note [Tarski, 1929b]. In 1933, Tarski outlined a proof of the deduction theorem for a particular formalized theory (see [Tarski, 1933b, Theorem 2(a), p. 286]). The deduction theorem is often attributed to Herbrand, who published the result in [Herbrand, 1928].

Tarski defines for this subcategory nearly all of the metamathematical notions that he defined in [Tarski, 1930d]] for deductive systems generally, and in much the same way.[14] He also states without proof a dozen or so theorems. Among these are the first published statement of Lindenbaum's lemma (Theorem 12, p. 34), and the result that every (countable) set of sentences possesses a base (Theorem 17, p. 35) — a result that holds for the present subcategory but not generally, as we noted above.

2.3 The sentential calculus

Another subcategory of deductive systems — the sentential calculus — is investigated in Lukasiewicz and Tarski's "Investigations into the sentential calculus" [Tarski, 1930c]. The authors remark that

> "as the simplest deductive discipline, the sentential calculus is particularly suitable for metamathematical investigations. It can be regarded as a laboratory in which metamathematical methods can be discovered and mathematical concepts constructed which can then be carried over to more complicated mathematical systems." (p. 59)

[Tarski, 1930c] is "a compilation of theorems and concepts belonging to five different persons" (p. 38, fn ‡), namely, Lukasiewicz, Tarski, Lindenbaum, Sobocinski, and Wajsberg — all members of the Warsaw School of Logic.

At the outset of [Tarski, 1930c], Lukasiewicz and Tarski refer the reader to the conceptual apparatus and notation developed in [Tarski, 1930b]. The concepts of *sentential variable*, *conditional* and *negation* are taken as primitive. The metamathematical expressions '$c(x, y)$' and '$n(x)$' denote respectively the conditional with antecedent x and consequent y, and the negation of x. The set S of all sentences is defined as follows:

> "The *set S of all sentences* is the intersection of all those sets which contain all sentential variables (elementary sentences) and are closed under the operations of forming implications [conditionals] and negations." (p. 39)

Lukasiewicz and Tarski go on to define the notion of consequence in terms of *substitution* and *detachment* (or modus ponens).

> "The *set of consequences* $Cn(X)$ *of the set* X *of sentences* is the intersection of all those sets which include the set $X \subseteq S$ and are closed under the operations of substitution and detachment." (p. 40)

From this definition it follows that S and $Cn(X)$ satisfy the axioms 1–5 of [Tarski, 1930b] (see above). Now consider X such that $X \subseteq S$ and X is closed under consequence (that is, $Cn(X) = X$). Then by Tarski's characterization of a deductive

[14] The only exceptions are the notions of *decision domain* and *cardinal degree of completeness.*

system, X is a deductive system. It is these deductive systems that Lukasiewicz and Tarski investigate in [Tarski, 1930c].

Lukasiewicz and Tarski identify two methods of constructing such a deductive system. One is the familiar *axiomatic method*, according to which

> "an arbitrary, usually finite, set X of sentences — an *axiom system* — is given, and the set $Cn(X)$, i.e. the smallest deductive system over X, is formed." (p. 40)

The second method is the *matrix method*. The following definition of a matrix is due to Tarski:

> "A *(logical) matrix* is an ordered quadruple $M = [A, B, f, g]$ which consists of two disjoint sets (with elements of any kind whatever) A and B, a function f of two variables and a function g of one variable, where the two functions are defined for all elements of the set $A \cup B$ and take as values elements of $A \cup B$ exclusively." (p. 41)[15]

The elements of the set B are called the designated elements (following Bernays). For an example of a matrix, set A, B, f, g as follows: $A = \{0\}, B = \{1\}, f(0,0) = f(0,1) = f(1,1) = 1, f(1,0) = 0, g(0) = 1, g(1) = 0$. This is the matrix associated with the classical sentential calculus (and the designated value is 1). The idea is that the functions f and g correspond to the syntactic operations of forming conditionals and negations.[16] This correspondence is made precise by Tarski's recursive definition of a value function:

> "The function h is called a *value function* of the matrix $M = [A, B, f, g]$ if it satisfies the following conditions:
>
> 1. the function h is defined for every $x \in S$;
> 2. if x is a sentential variable, then $h(x) \in A \cup B$;
> 3. if $x \in S$ and $y \in S$, then $h(c(x, y)) = f(h(x), h(y))$;
> 4. if $x \in S$ then $h(n(x)) = g(h(x))$." (p. 41)

Given the notion of a value function, Tarski defines the notion of satisfaction of a sentence by a matrix:

[15] I have replaced Tarski's symbol '+' for set-theoretical union by the more familiar '∪'.

[16] The functions f and g may be captured by the these tables for '→' and '¬':

x	y	$x \to y$
1	1	1
1	0	0
0	1	1
0	0	1

x	$\neg x$
1	0
0	1

Replacing '1' and '0' by 'T' and 'F' yields the familiar truth tables (where T is the designated value).

> "The sentence x is *satisfied (or verified) by the matrix* $M = [A, B, f, g]$ ... if the formula $h(x) \in B$ holds for every value function h of this matrix." (p. 41)

In order to construct a deductive system of the sentential calculus we proceed as follows: we set up a matrix and consider all those sentences satisfied by it.[17]

In [Tarski, 1930c], the matrix method is used to construct not only the classical (two-valued) system of the sentential calculus, but also an entire class of many-valued systems. The definition of the classical system L of the sentential calculus has already been anticipated:

> "The *ordinary [classical] system L of the sentential calculus* is the set of all sentences which are satisfied by the matrix $M = [A, B, f, g]$ where $A = \{0\}, B = \{1\}$ and the functions f and g are defined by the formulas: $f(0,0) = f(0,1) = f(1,1) = 1, f(1,0) = 0, g(0) = 1, g(1) = 0$." (p. 42)

Lukasiewicz and Tarski observe that the consistency and completeness of L follows easily from this definition. They also define L by the axiomatic method, presenting an axiom system due to Lukasiewicz, namely the set $\{(p \to q) \to ((q \to r) \to (p \to r)),\ (\neg p \to p) \to p,\ p \to (\neg p \to q)\}$.

Tarski proved the following noteworthy theorem about L:

> The system L, as well as every axiomatizable system of the sentential calculus which contains the sentences '$p \to (q \to p)$' and '$p \to (q \to ((p \to (q \to r)) \to r))$', possesses a basis consisting of a single sentence. (Theorem 8, p. 44.)

A generalization of this theorem is also stated:

> The system L, as well as every axiomatizable system of the sentential calculus which contains the sentences '$p \to (q \to p)$' and '$p \to (q \to ((p \to (q \to r)) \to r))$', possesses for every natural number m a basis containing exactly m elements. (Theorem 10, p. 45.)[18]

Lukasiewicz and Tarski also use the matrix method to define a class of many-valued systems founded by Lukasiewicz — the so-called "n-valued systems":

[17] This procedure rests on the following theorem:
If M is a normal matrix, then the set of all sentences satisfied by M is a deductive system, where a matrix is normal if the formulas $x \in B$ and $y \in A$ always imply $f(x, y) \in A$.

The generality of the matrix method is established by a converse of this theorem, due to Lindenbaum. Lindenbaum's theorem runs as follows:

For every deductive system X there exists a normal matrix $M = [A, B, f, g]$, with an at most denumerable set $A \cup B$, which satisfies exactly the sentences of X. (See [Tarski, 1930c, pp. 41–42].)

[18] The proof of the theorem for L is credited to Sobocinski, and the generalizations to other systems to Tarski.

Tarski also proved the following theorem, which provides a contrast between L and other systems of the sentential calculus: "Theorem 11. For every natural number m, systems of the sentential calculus exist every basis of which contains exactly m elements." [Tarski, 1930c, p. 45]

> "The *n-valued system* L_n *of the sentential calculus* (where n is a natural number or $n = \aleph_0$) is the set of all sentences which are satisfied by the matrix $M = [A, B, f, g]$ where, in the case $n = 1$ the set A is null, in the case $1 < n < \aleph_0$ A consists of all fractions of the form $k/n-1$ for $0 \leq k < n-1$, and in the case $n = \aleph_0$ it consists of all fractions k/l for $0 \leq k < l$; further the set B is equal to $\{1\}$ and the functions f and g are defined by the formulas: $f(x, y) = \min(1, 1-x+y), g(x) = 1-x$." (pp. 47–8)[19]

Lukasiewicz and Tarski report a number of significant results about these systems. They include the following:

(i) For every $n, 1 \leq n < \aleph_0, L_n$ is axiomatizable. (Theorem 22, p. 49).[20]

(ii) Let $M = [A, B, f, g]$ be a normal matrix [see fn. 17] in which the set $A \cup B$ is finite. If the sentences '$(p \to q) \to ((q \to r) \to (p \to r))$', $(q \to r) \to ((p \to q) \to (p \to r)), (p \to q) \to (\neg q \to \neg p), \neg q \to ((p \to q) \to \neg p)$ are satisfied by M, then the set of all sentences satisfied by M is axiomatizable. (Theorem 24, p. 50. Tarski notes that Theorem 24 is a generalization of Theorem 22.)

(iii) Every system L_n, where $2 \leq n < \aleph_0$, possesses, for every natural number m (and in particular for $m = 1$), a basis which has exactly m elements. (Theorem 26, p. 50.)

In comparing the axiomatic method and the matrix method, Tarski writes:

> "Each of the two methods has its advantages and disadvantages. Systems constructed by means of the axiomatic method are easier to investigate regarding their axiomatizability, but systems generated by matrices are easier to test for completeness and consistency." (p. 42)

[19] For example, consider the case $n = 3$. Then $A = \{0, \frac{1}{2}\}$ and $B = \{1\}$. The functions f and g may be captured by the following tables for $\to$ and $\neg$ respectively:

p	q	$p \to q$
1	1	1
1	$\frac{1}{2}$	$\frac{1}{2}$
1	0	0
$\frac{1}{2}$	1	1
$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{2}$	0	$\frac{1}{2}$
0	1	1
0	$\frac{1}{2}$	1
0	0	1

p	$\neg p$
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

Given values m for p and n for q, the value of $p \to q$ is $\min(1, 1-m+n)$, and the value of $\neg p$ is $1-m$.

In the given definition, A is the set of all proper fractions. Lindenbaum showed that any other denumerably infinite subset of the interval $\langle 0, 1\rangle$ will serve as well.

[20] This theorem was proved by Wajsberg for $n = 3$ and for all n for which $n-1$ is a prime number, and later extended to all natural numbers by Lindenbaum.

Tarski was keenly aware of the value of the matrix method as an alternative to the axiomatic method. Blok and Pigozzi point out that although Tarski was not the first to use matrices to construct deductive systems (or 'logics'), he was responsible for recognizing the importance of matrices for general metamathematics. They continue:

> "It is difficult to overestimate their importance in this regard. They provided a means of defining logics that were not *a priori* finitely axiomatizable, and this led naturally to the question of the existence of logics that are not finitely axiomatizable. They were used first by Bernays and then extensively by Lukasiewicz to establish the independence of various axiom systems. Finally, they provided a natural way of defining the equivalence of deductive systems with different primitive connectives. All three of these topics proved to be a rich source of problems." [Blok and Pigozzi, 1988, p. 421])

In addition, there is fruitful interplay between the two methods. Corcoran observes that [Tarski, 1930c]

> "contains results achieved by interrelating axiomatic and matrix methods for defining sentential calculi, an interrelation that clearly foreshadows what will later be seen as the interrelation of proof-theoretic (syntactic) and model-theoretic (semantic) methods." [Tarski, 1983a, p. xix]

2.4 The algebraization of logic

Tarski's paper "Foundations of the Calculus of Systems" [Tarski, 1935+1936] extends the metamathematical investigations of [Tarski, 1930b; 1930c] and [Tarski, 1930d]. Like [Tarski, 1930b], the paper considers only those theories "whose construction presupposes a logical basis of a greater or lesser extent, and at least the whole sentential calculus" (p. 342). But the axiom system of [Tarski, 1930b] is replaced by a "simpler", "more natural" one [Tarski, 1935+1936, p. 342]. Let 'L' denote the smallest deductive system defined by Axioms 1–10 of [Tarski, 1930b] — that is, the system $C(\varnothing)$, the set of consequences of the empty set. In the axiom system of [Tarski, 1935+1936], the symbol 'L' is taken as a primitive expression instead of '$Cn(X)$' — the concept of *consequence* is no longer a primitive notion.

The new axiom system consists of the following five axioms:

Axiom 1. $0 < \overline{\overline{S}} \leq \aleph_0$ (where S is the set of meaningful sentences)

Axiom 2. If $x, y \in S$, then $\bar{x} \in S$ and $x \rightarrow y \in S$ (where '$\bar{x}$' is the negation of x).

Axiom 3. $L \subseteq S$.

Axiom 4. If $x, y, z \in S$, then $(\bar{x} \to x) \to x \in L$, $x \to (\bar{x} \to y) \in L$ and $(x \to y) \to ((y \to z) \to (x \to z)) \in L$.

Axiom 5. If $x, x \to y \in L$ (where $y \in S$), then $y \in L$.

L can be interpreted as the set of all logically valid sentences. It is straightforward to define the concept of consequence, once we have defined the concepts of the *sum* and the *product* of sentences:

Definition. $x + y = \bar{x} \to y$, $x.y = \overline{(x \to \bar{y})}$ for all $x, y \in S$.[21]

Tarski goes on to show that this system of axioms and primitive concepts is equivalent to that of [Tarski, 1930b].

At the heart of [Tarski, 1935+1936] are two deductive systems, *each presented as an interpretation of Boolean algebra.* Tarski writes:

> "[T]wo calculi can be constructed which are very useful in metamathematical investigations, namely the *calculus of sentences* and the *calculus of deductive systems*; the first is a complete and the second a partial interpretation of the formal system which is usually called the *algebra of logic* or *Boolean algebra*." (p. 347)

Tarski identifies eight primitive concepts: 'B' is 'the universe of discourse', '$<$' denotes the relation of inclusion, '$=$' denotes identity, '$+$' and '$.$' denote respectively the *sum of* and *product* operations, '0' denotes the zero (or empty) element, '1' denotes the unit (or universal) element, and '$\bar{x}$' denotes the complement of the element x. There are seven postulates that suffice for Boolean algebra:

Postulate I. (a) If $x \in B$, then $x < x$; (b) if $x, y, z \in B$, $x < y$ and $y < z$, then $x < z$.

Postulate II. If $x, y \in B$, then $x = y$ if and only if both $x < y$ and $y < x$.

Postulate III. If $x, y \in B$, then (a) $x + y \in B$; (b) $x < x + y$ and $y < x + y$; (c) if $z \in B$, $x < z$ and $y < z$, then $x + y < z$.

Postulate IV. If $x, y \in B$, then (a) $x.y \in B$; (b) $x.y < x$ and $x.y < y$; (c) if $z \in B$, $z < x$ and $z < y$, then $z < x.y$.

Postulate V. If $x, y, z \in B$, then (a) $x.(y + z) = x.y + x.z$ and (b) $x + (y.z) = (x + y).(x + z)$.

[21] In order to define the concept of consequence, the definition of the sum and the product of sentences is extended by recursion to an arbitrary finite number of disjuncts and conjuncts:

Definition. $\sum_{i=1}^{n} x_i = \prod_{i=1}^{n} x_i = x_1$ if $n = 1$ and $x_1 \in S$; $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n-1} x_i + x_n$ and $\prod_{i=1}^{n} x_i = \prod_{i=1}^{n-1} x_i.x_n$, if n is an arbitrary natural number > 1 and $x_1, x_2, \ldots, x_n \in S$.

The definition of *consequence* is as follows:

Definition For an arbitrary set $X \subseteq S$ the set $Cn(X)$ consists of those, and only those, sentences $y \in S$ which satisfy the following condition: either $y \in L$, or there exist sentences $x_1, x_2, \ldots, x_n \in X$ such that $(\prod_{i=1}^{n} x_i \to y) \in L$.

Postulate VI. (a) $0, 1 \in B$; (b) if $x \in B$, then $0 < x$ and $x < 1$.

Postulate VII. If $x \in B$, then (a) $\bar{x} \in B$, (b) $x.\bar{x} = 0$, and (c) $x + \bar{x} = 1$.

These are the axioms for the general, abstract theory of Boolean algebra. Boolean algebra is an abstraction from Boolean set algebra, in which B is the universe of sets, '$<$' denotes set-theoretical inclusion, '$+$' and '.' denote respectively set-theoretical union and intersection, 0 and 1 are the empty set and universal set respectively, and '$\bar{x}$' is the set-theoretical complement of x.

Tarski turns first to the *calculus of sentences* (which he also calls "the sentential algorithm" to avoid confusion with the expression 'sentential calculus'). Its universe of discourse is S, and two relations are defined on S: the relation of *implication*, denoted by '$\supset$', and the relation of *equivalence*, denoted by '$\equiv$'.

Definition. (a) $x \supset y$ if and only if $x, y \in S$ and $x \to y \in L$; (b) $x \equiv y$ if and only if both $x \supset y$ and $y \supset x$.

Now Tarski states the following (easily checked) theorem:

Theorem. Postulates I–VII are satisfied under the following replacements: the symbols 'B', '$<$', and '$=$' are replaced respectively by 'S', '$\supset$', and '$\equiv$'; and '0' is replaced by the variable 'u' and '1' by the variable 'v', where it is assumed that $u \in S, \bar{u} \in L$, and $v \in L$. [Tarski, 1935+1936, Theorem 4, p. 348].

This theorem shows that the Boolean algebra of Postulates I–VII can be derived from the Axioms 2–5.[22] And the result holds in the other direction. Tarski sums up:

> "We can thus assert that Axs. 2–5 form a system of statements which is equivalent to the system of postulates for the ordinary algebra of logic" [Tarski, 1935+1936, p. 349].

In making fully precise the relation between the system L and Boolean algebra, Tarski sets the stage for algebraic logic. The calculus of systems can now be seen as an interpretation of Boolean algebra. As Blok and Pigozzi put it, "... here for the first time can be found all the essential features of modern algebraic logic" [Blok and Pigozzi, 1988, pp. 45–6]; and as Vaught puts it, "... Tarski introduced here the now well-known Boolean algebra $B(L)$ canonically associated with a theory L; and he initiated the still continuing study of the algebras $B(L)$" [Vaught, 1986, p. 873]. Tarski observes that the present interpretation of Boolean algebra is easily modified so as to avoid the replacement of identity by another equivalence relation (viz. $\equiv$). This modification yields a very early instance of an important and now-familiar category of Boolean algebras: the Lindenbaum-Tarski algebras.[23]

[22] Axiom I plays no part in the proof of the theorem.

[23] Tarski details the modification in [Tarski, 1935+1936, fn. 1, p. 349]. Given any $x \in S$, consider its equivalence class $X \subseteq S$ — viz., the set of all sentences y such that $y \equiv x$. For these equivalence classes the relations and operations $\supset, +, .$, etc. are defined in the appropriate way; for example, '$X \supset Y$' says that $x \supset y$ for all $x \in X$ and $y \in Y$. 0 is the set of all sentences $x \in S$

Despite the lasting significance of the calculus of systems, the "chief subject" of [Tarski, 1935+1936] is another calculus, the *calculus of deductive systems*, or for short, the *calculus of systems*. Recall that a deductive system is a system closed under consequence. Given the definition of *consequence* (see fn. 21), the notion of a deductive system can be characterized as follows:

> X is a deductive system iff $L \subseteq X \subseteq S$ and if $x,\ x \to y \in X$ (where $y \in S$), then $y \in X$.

In the calculus of systems, there is a correlate of each primitive concept of Boolean algebra. The universe of discourse is the class of deductive systems. L is the zero system and S is the unit system. Inclusion and identity are interpreted in the ordinary set-theoretical way — and, similarly, the product of systems is interpreted as set-theoretical intersection. However, the addition of systems is not set-theoretical union — the union of two systems does not in general yield a new deductive system. Instead, the addition of systems X and Y, denoted by '$X \dot{+} Y$', is defined as $Cn(X \cup Y)$. Similarly, complementation is not set-theoretical complementation; instead, the complement $\bar{X}$ of system X is the sum of all systems Y disjoint with X (i.e. such that $X \cap Y = L$). The calculus of systems is a *partial interpretation of Boolean algebra*. Consider the postulates I–VII, and replace the variables 'x', 'y', 'z' respectively by 'X', 'Y', 'Z', and the constants 'B', '$<$', '$+$', '0' and '1' respectively by 'D', '$\subseteq$', '$\dot{+}$', 'L' and 'S'. Then every postulate except VIIc is satisfied. And the following consequence of VIIc is also satisfied:

VIId. If $X, Y \subseteq D$, and $X \cap Y = L$, then $Y \subseteq \bar{X}$. (See Theorem 6, p. 351.)

So the law of excluded middle — $X \dot{+} \bar{X} = S$ — fails in the calculus of systems, and it is here that the calculus of systems differs essentially from the ordinary calculus of classes or sets. Tarski observes:

> "The formal resemblance of the calculus of systems to the intuitionistic sentential calculus of Heyting is striking: we might say that the formal relation of the calculus of systems to the ordinary calculus of classes is exactly the same as the relation of Heyting's sentential calculus to the ordinary sentential calculus" (p. 352)

In other words, the system of Postulates I–VI and VIIa,b,d is

> "a sufficient basis for a system of the algebra of logic which has the intuitionistic calculus as one of its interpretations" (*ibid.*)[24]

such that $\bar{x} \in L$, and 1 is the set L. (Observe that the identity relation is left unchanged.) The set of equivalence classes is a Boolean algebra — the Lindenbaum-Tarski algebra determined by L.

[24] For a more detailed elaboration of these remarks, Tarski refers the reader to his paper [Tarski, 1938b] (in particular section 5) and Stone [1937–38]. (The references are given in [Tarski, 1983a, p. 352, fn. †], where a typographical error appears — 'X and VII" should read "XVII".)

Tarski goes on to state a large number of results about the calculus of systems that are important both for metamathematics and model theory.[25] In section 3, Tarski investigates axiomatizable and non-axiomatizable systems. In section 4, he studies irreducible[26] and complete[27] systems, and shows how to characterize certain classes of systems in terms of cardinality and structure. In section 5 and the appendix, Tarski applies these general results to particular deductive theories.[28]

We have seen, then, that Tarski's metamathematical investigations in [Tarski, 1935+1936] are carried out in the framework of Boolean algebra; they "do not transcend the boundary of the algebra of logic" (p. 350). In a footnote later added to [Tarski, 1935+1936], Tarski observes:

> "In fact the calculus of deductive systems outlined in this paper proves to coincide with what was somewhat later developed as the calculus of Boolean-algebraic ideals. . . "[29]

Subsequently, it was the study of Boolean algebras that took center stage for Tarski, not the calculus of systems. In his paper [Tarski, 1938b], for example, Tarski refers to "*general metamathematics*, i.e. the *theory of deductive systems*" as "[a]nother important realization of Boolean algebra" [Tarski, 1938b, p. 454].[30] For more on Tarski and algebraic logic, see Section 6 below.

2.5 Metamathematics and models

In Chapter VI of [Tarski, 1941a], Tarski provides a clear, textbook account of the methodology of the deductive sciences which is strikingly semantic in character. As we have just seen, the semantic concepts of *interpretation* and *realization* figure in [Tarski, 1935+1936] — but in [Tarski, 1941a] they are at the heart of Tarski's presentation of metamathematics.

[25] For a summary of some of these results, see Blok and Pigozzi [1988, p. 44], and Vaught [1986, p. 873].

[26] X is an irreducible system iff $X \neq L$ and for every deductive system Y such that $Y \subseteq X, Y = L$ or $Y = X$.

[27] X is a complete system iff X is a deductive system, $X \neq S$ and for all deductive systems Y such that $X \subseteq Y, Y = X$ or $Y = S$.

[28] These theories include the theory of dense orders, the theory of discrete orders, the theory of identity, the theory of atomistic Boolean algebras, the general theory of order, and the general theory of binary relations.

[29] Fn.1, p. 350. Given a Boolean algebra characterized by Postulates I–VII, an *ideal* is a non-empty subset K of B such that (1) if $x \in K$ and $y \in K$, then $x + y \in K$, and (2) if $x \in K$ and $y \in B$, then $x.y \in K$.

Tarski refers the reader to Stone [1936; 1937], with which there is significant overlap or connection.

[30] As Tarski observes (in fn †, p. 352 of [Tarski, 1935+1936]), [Tarski, 1938b] contains "a more exact and detailed formulation" of the relation mentioned above between the calculus of systems and the intuitionistic sentential calculus. But [Tarski, 1938b] proceeds in terms of Boolean algebra, and not at all in terms of the calculus of systems. It is only at the very end of the paper that Tarski remarks that the results shown to hold for the formal system of Boolean algebra also hold for every realization of this system, such as the theory of fields of sets and the calculus of systems. (For a summary of results in [Tarski, 1938b], see Blok and Pigozzi [1988, pp. 46–7]).

Tarski starts out with a concrete example of a simple deductive theory, the theory of the congruence of line segments. Let the variables x, y, z range over line segments. There are two primitive terms 'S' and '$\cong$', where 'S' denotes the set of all line segments, and '$\cong$' denotes the relation of congruence. There are two axioms:

Axiom I. For all $x \in S, x \cong x$.

Axiom II. For all $x, y, z \in S$, if $x \cong z$ and $y \cong z$, then $x \cong y$.

Tarski goes on to derive two theorems: the first says that the congruence relation is symmetric, the second that the relation is transitive.

Tarski observes that though our knowledge of segments and congruence goes a long way beyond the axioms, this additional knowledge plays no role in the construction of the theory.

> "In particular, in deriving theorems from the axioms, we make no use whatsoever of this knowledge, and behave as though we did not understand the content of the concepts involved in our considerations, and as if we knew nothing about them that had not been expressly asserted in the axioms. We disregard, as it is commonly put, the meaning of the primitive terms adopted by us, and direct our attention exclusively to the form of the axioms in which these terms occur." (p. 122)

Accordingly, we can abstract away from our particular theory by replacing 'S' by a variable 'K' that ranges over all classes, and '$\cong$' by a variable 'R' that ranges over all (2-place) relations, so as to obtain:

Axiom I′. For all $x \in K, xRx$, and

Axiom II′ For all $x, y, z \in K$, if xRz and yRz, then xRy.

Now the statements of the theory are logical statements; for example, Axiom I' says that the relation R is reflexive on K. And any theorem about congruent segments is now correlated with a general law in the domain of logic (for example, that any relation R for which the generalized Axioms I' and II' hold is symmetric).

At this point, Tarski introduces the notion of a *model* (or *realization*) *of the axiom system.* If, in a set K, a relation R is reflexive and has the property expressed by Axiom II′, then K and R together form a model of the axiom system I′ and II′. One model is, of course, formed by the class of segments and the congruence relation — but this is not in any way privileged. Another model of the axiom system is provided by the universal class and the relation of identity. And another is formed by the set of all numbers — or any set of numbers — and the relation given by 'the difference between numbers x and y is an integer'. Every model of the axiom theory will also satisfy all theorems deduced from these axioms.

These general facts, Tarski says, "have many interesting applications in methodological researches" [Tarski, 1941a, p. 124]. For example, it may be proved from these facts that certain sentences cannot be deduced from the axiom system I and II. Consider the following sentence:

A. There exist two elements x and y of the set S for which it is not the case that $x \cong y$. (There exist two segments that are not congruent).

Though A is true, attempts to prove it from the axioms fail. Can we *show* that the attempt must fail? We can, by applying what Tarski here calls the *method of proof by interpretation*.[31]

> "If sentence A could be proved on the basis of our axiom system, then, as we know, every model of this system would satisfy that sentence; if, therefore, we succeed in indicating such a model of the axiom system which will not satisfy Sentence A, we shall prove thereby that this sentence cannot be deduced from Axioms I and II." (p. 125)

And such a model is easy to find — for example, consider the model formed by the set of integers, together with the relation 'the difference between numbers x and y is an integer'.

Abstracting away from any particular axiom system and its models, Tarski goes on to state "a general law from the domain of the methodology of the deductive sciences".[32]

> "Every theorem of a given deductive theory is satisfied by any model of the axiom system of this theory; and moreover, to every theorem there corresponds a general statement which can be formulated and proved within the framework of logic and which establishes the fact that the theorem in question is satisfied by any such model." (p. 127)

This law — a version of what Tarski calls the *law of deduction* — has "tremendous practical importance" (p. 127). Given an axiom system of a given theory, we will often find that the constants of another deductive theory form a model of that axiom system — that is, we find an interpretation of the axiom system of the original theory within the other theory.[33] And the validity of the theorems of the

[31] As we shall see, the method here is based on Tarski's definition of *logical consequence* in terms of models — see Section 3.4 below. The definition runs as follows:

The sentence X *follows logically* from the sentences of the class K if and only if every model of the class K is also a model of the sentence X. [Tarski, 1936b, p. 417].

We show that A does not follow logically from Axioms I and II by constructing a model of the axiom system which is not a model of A.

[32] In leading up to the statement of this law, Tarski makes the simplifying assumption that logic is the only theory preceding the given theory (see footnote 6 above).

[33] For example, the axiom system of arithmetic may be interpreted within geometry (enabling us to obtain a visual images of various facts in the field of arithmetic); and conversely, the axiom system of geometry has an interpretation within arithmetic (giving rise to analytic geometry, in which geometrical facts are investigated with the help of arithmetical or algebraic methods). Further, arithmetic is interpretable within logic (and hence, so is geometry). See [Tarski, 1941a, pp. 129–130].

first theory carries over to those of the second: '*All theorems proved on the basis of a given axiom system remain valid for any interpretation of the system.*" (p. 128) So in proving a theorem of the original theory we prove any number of others:

> "Every proof within a deductive theory contains — potentially, so to speak — an unlimited number of other analogous proofs." (p. 128)

These facts demonstrate "the great value of the deductive method from the point of view of economy of human thought" (*ibid*). They are also of far-reaching theoretical importance, since

> "they establish a foundation for various arguments and researches within the methodology of deductive sciences" (p. 128).

For example, the law of deduction is the theoretical basis for all proofs by interpretation.

In the final section of Chapter VI, Tarski stresses the broad scope of the methodology of the deductive sciences. As we have seen throughout the present section, it goes beyond the study of methods used in the construction of deductive sciences, and incorporates the study of deductive theories as *wholes*; consider, for example, investigations into the consistency and completeness of theories. Better, says Tarski, to call it "metalogic and metamathematics". Rather than the study of methods, it is become a general science of deductive sciences, requiring its own precise methods. Tarski writes:

> "Methodology has become like those sciences which constitute its own subject matter - it has assumed the form of a deductive discipline". (p. 140)

2.6 Some historical remarks

Understood in this suitably broad way, the methodology of the deductive sciences is, Tarski says, "a very young discipline".

> "Its intensive development began only twenty years ago — simultaneously (and, as it seems, independently) in two different centers: Gottingen under the influence of D. Hilbert and P. Bernays, and Warsaw, where S. Leśniewski and J. Lukasiewicz, among others, worked." [Tarski, 1941a, p. 19]

Its roots, however, go back a very long way. We can find an analysis of the deductive sciences in Aristotle's *Posterior Analytics*.[34] Much more recent, but still of

[34] [Tarski, 1935]. For example, according to Aristotle every demonstrative science has three elements: (1) what is posited, (2) the axioms, and (3) the attributes. Consider for illustration the case of geometry. In geometry, points and lines are posited, and their attributes are demonstrated "by means of the axioms and from previous conclusions as premisses" (76b9–10, Bk.1, Ch.10, *Posterior Analytics*, McKeon [1941, p. 24]).

historical significance for the methodology of the deductive sciences, are Bolzano's *Wissenschaftslehre* [1837], Pasch [1882], Hilbert's *Grundlagen der Geometrie*, and Veblen and Young [1910].

Applications of the deductive method have a very long history too. Of Euclid's *Elements*, Tarski writes:

> "For 2200 years, mathematicians have seen in Euclid's work the ideal and prototype of scientific exactitude." [Tarski, 1941a, p. 120, fn. 1]

Tarski continues:

> "An essential progress in this field occurred only in the last 50 years, in the course of which the foundations of the basic mathematical disciplines of geometry and arithmetic were laid in accordance with all requirements of the present-day methodology of mathematics." (p. 120, fn. 1)

Tarski mentions in particular Peano's *Formulaire de Mathematiques*, and Hilbert's *Grundlagen der Geometrie*. As they stood, Peano's and Hilbert's theories failed to meet one of the requirements of methodology, that the deductive theory be formalized. The development of formalized deductive theories started with Frege:

> "The first attempts to the present the deductive theories in a formalized form are due to Frege.... A very high level in the process of formalization was achieved in the works of the late Polish logician S. Leśniewski...." (p. 133)

It is clear that Tarski's work in metamathematics continued a long tradition, and that even the "young discipline" of the methodology of the deductive sciences was not original with Tarski. But it is also clear that Tarski took metamathematics to a new level. Corcoran writes:

> "The modern conception, and perhaps the very existence, of the methodology of deductive science as a separate science and in particular as a formalizable deductive science, is largely due to Tarski in that it was he who collected, analyzed, and codified emerging concepts and results, and it was he who most clearly and forcefully articulated and defended the possibility of such a science." (Corcoran, editor's introduction to [Tarski, 1983a, p. xvi]

3 FORMAL DEFINABILITY

As we have seen, one of the tasks of the methodology of the deductive sciences is to "make precise the meaning ... of important metamathematical concepts" [1930e]. In a series of papers published in the 1930s, Tarski carried out this project for a

number of fundamental concepts — most notably *definability*, *truth*, and *logical consequence*. Tarski constructed precise definitions of each of these concepts.

Tarski distinguished two kinds of definition:

> "The words 'define', 'definable', etc., are used in two distinct senses: in the first sense it is a question of a *formal* relation of certain expressions to other expressions of a theory. . . ; in the second sense of a *semantical* relation between objects and expressions." [Tarski, 1934–35, fn. 1, p. 386]

Broadly speaking, formal definitions involve only 'syntactic' notions, such as derivability, while semantic definitions involve word-world relations. In [Tarski, 1934], Tarski provided a formal definition of *definability*. In his papers [Tarski, 1931a; Tarski, 1933a] and [Tarski, 1936b], Tarski constructed semantic definitions for *definability*, *truth*, and *logical consequence* respectively. In this section, we will explore Tarski's formal treatment of definability, and its ramifications. In Section IV below, we turn to Tarski's semantic treatment of semantic definability, truth, and logical consequence.

The formal sense of 'definability' is investigated in Tarski's paper "Some methodological investigations on the definability of concepts" [Tarski, 1934]. Tarski constructs a formal definition of definability that draws on earlier work of Padoa's [Padoa, 1900].

Padoa takes a deductive theory to be a system of *undefined symbols* (or uninterpreted primitive terms) and a system of *unproved propositions* (or axioms). Padoa's chief aim is to characterize an *irreducible* system of undefined symbols, that is, a system in which no undefined symbol can be defined in terms of the others:

> "We say that *the system of undefined symbols is irreducible with respect to the system of unproved propositions* when *no symbolic definition of any undefined symbol can be deduced from the system of unproved propositions*, that is, when we cannot deduce from the system a relation of the form $x = a$, where x is one of the undefined symbols and a is a sequence of other such symbols (and logical symbols)." (pp. 121–22)[35]

Padoa's definition of an irreducible system contains an informal characterization of formal definability that Tarski makes explicit and precise in [Tarski, 1934]. Tarski considers deductive theories with the simple theory of types as their logical basis, and directs his attention to the extra-logical constants, or *terms*, of a given

[35] As Padoa observes, there is a parallel definition of an *irreducible (or "absolutely independent") set of propositions* — viz, a set where no member can be derived from the other members. (Padoa points out that a method of proving this irreducibility is already well-known: we can say that a proposition is not a logical consequence of the others if there is an interpretation in which it is false, and the others are true. The method is now standard in independence proofs. Padoa cites Peano [1899, p. 30], which contains a proof of the absolute independence of the axioms of arithmetic.) Like Padoa, Tarski presses the analogy between the concept *derivable sentence* on the one hand and *definable concept* on the other.

deductive theory. Let 'a' be a term, and B any set of terms. A *definition of the term 'a' by means of the terms of the set B* is a sentence of the following form:

(I) $(x)[x = a \leftrightarrow \phi(x, b', b'', \ldots)]$

where '$\phi(x, b', b'', \ldots)$' is a sentential function with the sole free variable x, and in which no extra-logical constants other than $b', b'', \ldots$ of the set B occurs. Now let X be a set of sentences in which all terms of B occur. Tarski defines the notion of (formal) *definability* as follows:

Definition. The term 'a' *is definable by means of the terms of the set B on the basis of the set X of sentences* if a definition of the term 'a' by means of the terms of B is derivable from the sentences of X. (See [Tarski, 1934, p. 299].)

Observe that this notion of definability involves only formal or 'syntactic' notions, in particular those of *derivability*, *sentential function*, and *term*.

Having defined the notion of formal derivability, Tarski goes on to characterize and justify Padoa's method for determining the irreducibility of a system of undefined symbols (or, equivalently, the undefinability of a term by means of other terms). Padoa offers only a sketch of this method:

> "Let us assume that, after an *interpretation* of the system of undefined symbols that verifies the system of unproved propositions has been determined, all these propositions are still verified if we suitably change the meaning of the undefined symbol x *only*. Then, since the meaning of x is not *individualized* once we have chosen an *interpretation* of the *other* undefined symbols, we can assert that it is impossible to deduce a relation of the form $x = a$, where a is a sequence of other undefined symbols, from the unproved propositions." [Padoa, 1900, p. 122]

Tarski describes Padoa's method this way:

> "In order, by this method, to show that a term 'a' cannot be defined by means of the terms of a set B on the basis of a set X of sentences, it suffices to give two interpretations of all extra-logical constants which occur in the sentences of X, such that (1) in both interpretations all sentences of the set X are satisfied and (2) in both interpretations all sentences of the set B are given the same sense, but (3) the sense of 'a' undergoes a change." [Tarski, 1934, p. 300]

Tarski goes on to present some results "which provide a theoretical justification for the method of Padoa" (ibid). Consider again a term a, and a set B of terms $b', b'', \ldots$, where a is not a member of B. Let $c', c'', \ldots$ be the terms other than $a, b', b'', \ldots$ that occur in the sentences of X. The conjunction of all the sentences of X is represented in the schematic form: '$\psi(a; b', b'', \ldots; c', c'', \ldots)$'. Suppose that in every sentence of X, variables are uniformly substituted for some or all terms, where it is assumed that none of these variables become bound. In

this way, sentential functions are formed. The conjunction of all these sentential functions is represented schematically by making the same substitutions in '$\psi(a; b', b'', \ldots; c', c'', \ldots)$'. Tarski proves the following theorems:

Theorem. The term 'a' is definable by means of the terms of the set B on the basis of the set X of sentences if and only if the formula

II. $\forall x(x = a \leftrightarrow \exists z', z'', \ldots \psi(x; b', b'', \ldots; z', z'', \ldots))$

is derivable from the sentences of X. (This is Theorem 1, in [Tarski, 1934, p. 301].)

Theorem. The term 'a' is definable by means of the terms of the set B on the basis of the set X of sentences if and only if the formula

III. $\forall x', x'', y', y'', \ldots, z', z'', \ldots t', t'', \ldots (\psi(x'; y', y'', \ldots; z', z'', \ldots)$
$\& \ \psi(x''; y', y'', \ldots; t', t'', \ldots) \rightarrow x' = x'')$

is derivable. (This is Theorem 2 in [Tarski, 1934, p. 303].)

Theorem. The term 'a' is not definable by means of the terms of the set B on the basis of the set X of sentences if and only if the formula

IV. $\exists x', x'', y', y'', \ldots, z', z'', \ldots t', t'', \ldots (\psi(x'; y', y'', \ldots; z', z'', \ldots)$
$\& \ \psi(x''; y', y'', \ldots; t', t'' \ldots) \ \& \ x' \neq x'')$

is consistent. (This is Theorem 3 in [Tarski, 1934, p. 304].)

Theorem 3 follows immediately from Theorem 2.[36]

Theorem 3 "constitutes the proper theoretical foundation for the method of Padoa" [Tarski, 1934, p. 305]. For suppose that we want to establish that the term 'a' is not definable by the means of the terms of the set B. We apply the following procedure (see [Tarski, 1934, p. 305]). Take a deductive system Y, either shown to be consistent or assumed to be so. Look for terms $\bar{a}, \bar{\bar{a}}, \bar{b'}, \bar{b''}, \ldots, \bar{c'}, \bar{c''}, \ldots, \bar{\bar{c'}}, \bar{\bar{c''}}, \ldots$ which satisfy the following three conditions:

(i) Replacing the terms $a, b', b'', \ldots c', c'', \ldots$ by $\bar{a}, \bar{b'}, \bar{b''}, \ldots, \bar{c'}, \bar{c''}, \ldots$ in all sentences of X yields sentences of Y.

(ii) Replacing the terms $b', b'', \ldots$ by the terms $\bar{b'}, \bar{b''}, \ldots$, and the terms $a, c', c'', \ldots$ by the terms $\bar{\bar{a}}, \bar{\bar{c'}}, \bar{\bar{c''}}, \ldots$ in all sentences of X yields sentences of Y.

(iii) The system Y contains the formula $\bar{a} \neq \bar{\bar{a}}$.

If these conditions are met, then the following conjunction belongs to Y:

$$\psi(\bar{a}; \bar{b'}, \bar{b''}, \ldots; \bar{c'}, \bar{c''}, \ldots) \ \& \ \psi(\bar{\bar{a}}; \bar{b'}, \bar{b''}, \ldots; \bar{\bar{c'}}, \bar{\bar{c''}}, \ldots) \ \& \ \bar{a} \neq \bar{\bar{a}}.$$

[36] To show that Theorem 3 follows from Theorem 2, Tarski observes that formula IV is equivalent to the negation of formula III, and applies the principle that a formula is not provable if and only if its negation is consistent.

Formula IV of Theorem 3 is an easy consequence of this conjunction, and so formula IV belongs to Y. And since Y is consistent, formula IV is consistent. So by Theorem 3, the term 'a' is not definable by means of the terms of B. And here is Padoa's method (somewhat extended, since the procedure accommodates the occurrence of terms in the sentences of X that are different from 'a' and the terms in B.)

Thus Tarski has shown that we are justified in using Padoa's method in the simple theory of types. (As Tarski remarks in a footnote subsequently added to [Tarski, 1934, p. 300, fn. 2], Beth later made a very significant advance, proving Theorems 2 and 3 for a much wider class of deductive theories — the theories of first-order logic.)[37] The method has significant practical importance. Suppose we are considering the primitive terms of a given deductive theory. By applying Padoa's method as many times as there are primitive terms, we may determine whether or not the terms are mutually independent (or in Padoa's terminology, whether or not they form an *irreducible system.*) If the terms are not mutually independent, then we may eliminate the unnecessary ones — and this may lead to a simplification of the axiom system.

To illustrate this, Tarski considers n-dimensional Euclidean geometry. We may take the terms 'equidistant' and 'lying between' as the sole primitive terms of an axiom system of geometry. (The term 'equidistant' is the *metrical primitive term* and 'lying between' is the *descriptive primitive term* of geometry.) Tarski announces that by Padoa's method it can be shown that the term 'equidistant' is not definable by means of the term 'lying between'. However, for 2 dimensions and higher, the term 'lying between' is definable by means of the term 'equidistant'[38] — and the definition may be constructed according to the formula II in Theorem 1 above. Accordingly, Padoa's method yields a simpler axiom system for n-dimensional geometry ($n > 1$) in which 'equidistant' is the only primitive term.

So the formally defined notion of definability is the key to Tarski's presentation and justification of Padoa's method. In the second half of [Tarski, 1934], Tarski takes the notion of definability in another direction, using it to characterize what he calls "the problem of the completeness of concepts". This requires the notion of an "essentially richer" set, which rests on the notion of definability. Let Y be a set of sentences, and let $X \subseteq Y$. Then Y is *essentially richer* than X if and only if there are terms in Y (i.e. occurring in the sentences of Y) that are not in X, and these terms are not definable by means of the terms in X (not even on the basis of Y). Now we can define the notion of a set of sentences being *complete* with respect to its terms.

Definition. A set X of sentences is *complete with respect to its terms* if it is impossible to construct a categorical set Y of sentences which is essentially richer than X with respect to specific terms. (See [Tarski, 1934, p. 311].)[39]

[37] [Beth, 1953].

[38] See Lindenbaum and Tarski [1927b]. Tarski credits Lindenbaum with the result that the two terms are independent of each other only in the case of one-dimensional geometry.

[39] Notice the restriction here to *categorical* sets of sentences, where a set is *categorical* if any

The problem of the completeness of concepts can now be stated: given a set of sentences, is it complete or incomplete with respect to its specific terms? For specific examples, Tarski again turns to geometry. Consider *descriptive one-dimensional geometry*, the geometry of points and subsets of a straight line in which the only primitive concept is that of *lying between*. Consider the (easily formulated) categorical system X_1 which contains 'lying between' as its only primitive term. Now we can extend X_1 to the full metrical geometry of the straight line by adding 'equidistant' as a primitive term. Call this extended system X_2. As we saw above, the concept of *equidistance* cannot be defined in terms of the concept of *lying between*. Since X_2 is categorical and essentially richer than X_1, it follows that X_1 is not complete with respect to its terms.

Now the system X_2, like X_1, is not complete with respect to its terms. But we can extend X_2 to a system that *is*. We extend X_2 to an essentially richer categorical set of sentences X_3 by adding two new primitive terms, say '0' and '1', together with an axiom saying that these terms denote two distinct points. It turns out that X_3 is complete with respect to its terms. (X_3 is, in fact, formally identical to the arithmetic of the real numbers.)[40]

As a whole, Tarski's paper [Tarski, 1934] may be regarded as an investigation into the concept of *concept*. The formal definition of *definable concept* is the core of Tarski's solution to two problems: the problem of the definability and mutual

---

two interpretations of it are isomorphic. Tarski expresses the intuitive significance of the notion of categoricity this way:

> "A non-categorical set of sentences (especially if it is used as an axiom system of a deductive theory) does not give the impression of a closed and organic unity and does not seem to determine precisely the meaning of the concepts contained in it." [Tarski, 1934, p. 311]

Suppose we were to drop this restriction, and say that a set X of sentences is complete with respect to its terms if and only if it is impossible to construct an essentially richer set Y. Tarski points out (pp. 308–9) that, given this unrestricted definition, there will be no complete sets (apart from some trivial cases). For suppose that X is consistent, and does not contain all extra-logical constants, or terms. Now add to X an arbitrary logically provable sentence that contains a term not occurring on the sentences of X. This new set Y is an essentially richer set than X. For, as Tarski puts it,

> "... it is obvious that the new extra-logical constant cannot be defined by means of the terms of X. In fact the only sentence of Y in which this constant occurs is logically provable, and thus true independently of the meaning of the specific terms contained in it." [Tarski, 1934, p. 309, fn. 1].

[40] In [Tarski, 1934], Tarski goes on to prove a general theorem concerning completeness with respect to terms. First we need the notion of a *monotransformable* set of sentences:

Definition. A set of sentences is *monotransformable* if for any two interpretations of this set there is at most one relation which establishes the isomorphism of the two interpretations.

(Compare and contrast the notion of a categorical set of sentences: a set of sentences is categorical if for any two interpretations of it there is at least one relation which establishes the isomorphism of the two interpretations.)

Tarski's theorem is as follows:

Theorem. Every monotransformable set of sentences is complete with respect to its terms. [Tarski, 1934, p. 314]).

independence of concepts, and the problem of the completeness of concepts. As we have seen, this definition turns on *formal* relations between expressions, such as derivability. No semantic word-world relations are involved. The reverse is true for Tarski's *semantic* definitions of definability, truth and logical consequence, to which we now turn.

4 SEMANTICS: DEFINABILITY, TRUTH, CONSEQUENCE, MODEL THEORY

In three celebrated papers — "On definable sets of real numbers" [Tarski, 1931a], "The concept of truth in formalized languages" [Tarski, 1933a], and "On the concept of logical consequence" [Tarski, 1936b] — Tarski constructed precise definitions of *definability*, *truth*, and *logical consequence*. We turn to these constructions in Sections 4.1, 4.2 and 4.3 below. As we will see, at the core of each definition is the semantic notion of *satisfaction of a sentential function*. These semantic definitions were important contributions to the development of model theory, and in 4.4 I make some historical remarks about model theory.

At the outset of [Tarski, 1931a], Tarski notes that mathematicians are extremely wary of the notion of definability: "their attitude is one of distrust and reserve" [Tarski, 1931a, p. 110]. Tarski finds this perfectly understandable: the semantical notion of definability is ambiguous and subject to well-known paradoxes, such as Richard's, König's, and Berry's.[41] Tarski makes similar remarks about truth:

> "For although the meaning of the term 'true sentence' in colloquial language seems to be quite clear and intelligible, all attempts to define this meaning more precisely have hitherto been fruitless, and many investigations in which this term has been used and which started with apparently evident premisses have often led to paradoxes and antinomies.... The concept of truth shares in this respect the fate of other analogous concepts in the domain of the semantics of language". [Tarski, 1933a, p. 152][42]

Small wonder, then, that semantical concepts have an "evil reputation". [Tarski, 1933a, p. 252]

In his papers on definability, truth and logical consequence, Tarski's aim is to find rigorous characterizations of these notions and thereby place them beyond suspicion and 'safe' for theorists to use. Prior to Tarski, it was generally held that these semantic notions were not the business of the mathematician or logician. Of the notion of definability, Tarski writes:

> "The distrust of mathematicians towards the notion in question is reinforced by the current opinion that this notion is outside the proper limits of mathematics altogether." [Tarski, 1931a, p. 110]

[41] See Richard [1905], König [1905], and for Berry's paradox, Russell [1908].

[42] For related remarks about the concept of logical consequence, see [Tarski, 1936b, p. 409].

Tarski sought to reverse "the current opinion". He identified a metamathematical task here regarding definability, that of

> "making its meaning more precise, of removing the confusions and misunderstandings connected with it, and of establishing its fundamental properties. . . ." [Tarski, 1931a, p. 110]

Tarski continues:

> "I believe I have found a general method which allows us to construct a rigorous metamathematical definition of this notion." [Tarski, 1931a, p. 111]

Tarski later applied the same method to construct his definitions of truth and logical consequence.

According to Tarski, the method allows the construction of definitions that are not only "formally correct" but also "materially adequate", in the sense that they "grasp the current meaning of the notion as it is known intuitively" (p. 129 VI). At the core of Tarski's method is the notion of *satisfaction of a sentential function*, which appears in print for the first time in [Tarski, 1931a], Tarski's paper on definable sets of real numbers. We will now focus on [Tarski, 1931a] and the notion of definability.

4.1 Semantic definability

In [Tarski, 1931a], Tarski restricts his analysis of the notion of definability to just one category of objects: sets of real numbers.

> "The problem set in this article belongs in principle to the type of problems which frequently occur in the course of mathematical investigations. Our interest is directed towards a term of which we can given an account that is more or less precise in its intuitive content, but the meaning of which has not at present been rigorously established, at least in mathematics. We then seek to construct a definition of this term which, while satisfying the requirements of methodological rigour, will also render adequately and precisely the actual meaning of the term. It was just such problems that the geometers solved when they established the meaning of the terms 'movement', 'line', 'surface', or 'dimension' for the first time. Here I present an analogous problem concerning the term 'definable set of real numbers'. (pp. 111–112)

Tarski works within the framework of the simple theory of types (where the variables of the first order range over individuals, variables of the second order range over sets of individuals, and so on). The notion of a sentential function is introduced recursively. There are four *primitive sentential functions* denoted by '$\in$', '$\nu(x)$', '$\mu(x, y)$', '$\sigma(x, y, z)$'. The last three are special to the theory of

real numbers, where the first-order variables range over reals, and '$\nu(x)$' says that $x = 1$, '$\mu(x, y)$' says that '$x \leq y$', and '$\sigma(x, y, z)$' says that $x + y = z$. So the object of Tarski's investigation is the system of reals with the primitive concepts 1, $\leq$, and $+$, or the system $\langle R, 1, \leq, + \rangle$ for short. There are five *fundamental operations* on sentential functions: negation, logical sum '+' (disjunction), logical product '.' (conjunction), universal quantification, and existential quantification. The set of all sentential functions is the smallest set which contains the primitive functions and is closed under the five operations.

Tarski explicitly distingishes system and metasystem:

> "For each deductive system it is possible to construct a particular science, namely the 'metasystem', in which the given system is subjected to investigation" (p. 116).

Tarski stresses that the notion of definability is *relative* — we can define, in the metasystem, the notion of definability *for the given system*. Definability is relative to a system — and if we ignore this characteristic, we will fall into paradox.[43] In the present case, expressions such as 'variable of the nth order' and 'sentential function' are terms of the metasystem, applying as they do to particular expressions of the system under investigation. We can also introduce into the metasystem arithmetical notions, like *real number*, *set of real numbers*, and so on. The metasystem now has the resources to define the following phrase:

> A finite sequence of objects satisfies a given sentential function.

In [Tarski, 1931a], Tarski makes no attempt to define this phrase precisely — a rigorous treatment is given in [Tarski, 1933a]. Instead, Tarski presents some simple illustrations. For example, the primitive sentential function '$\sigma(x, y, z)$' is satisfied by all sequences of three real numbers x, y, z such that $x = y + z$; the sentential function '$\nu(x).\nu(y).\mu(y, z)$' is satisfied by all sequences of three real numbers x, y, z such that $x = 1 = y \leq z$; and the sentential function '$\exists z \exists u(\sigma(x, y, z).\mu(u, z).\nu(u)$' is satisfied by sequences of two real numbers x and y where $x \geq y + 1$. Tarski notes that these examples indicate the possibility of establishing a 1-1 correspondence between the members of the sequence and the free variables in the sentential function. In the limit case, a sentential function without free variables — that

[43] On p. 119, Tarski observes that there are only denumerably many definable sets (because there are only denumerably many sentential functions), but non-denumerably many sets of numbers. So there exist undefinable sets. Tarski goes on:

> "More than that, it is known that, with every denumerable family F of sets of numbers, a uniquely determined set F^* can be correlated that does not belong to F; taking for F the family of all definable sets, we get F^*, an example of a *set of numbers defined in terms of the metasystem, but not definable in the system itself.*" (p. 119)

If we ignore the system/metasystem distinction, we will be landed in contradiction and paradox. (In a footnote added later, Tarski refers the reader to [Tarski, 1948b, pp. 108f] for a precise definition of F^*.)

is, a sentence — "is satisfied either by the empty sequence or by no sequence, according as the sentence is true or false" (p. 117).

Of special significance are sentential functions with one free variable, where instead of (unit) sequences, we can speak of *objects* satisfying the given function. We are led naturally to the notion of a definable set:

> "Consequently, a function which contains a variable of order 1 as its only free variable determines a certain set of individuals, which, in particular, may be a certain set of real numbers. The sets thus determined by sentential functions are precisely the *sets definable in the arithmetical system considered*." (p. 118)

Tarski now states the metamathematical definition of definable sets of real numbers as follows:

(M) A set X is a *definable set* if there is a sentential function which contains some variable of order 1 as its only free variable, and which satisfies the condition that, for every real number $x, x \in X$ if and only if x satisfies this function. (p. 118)

For example, the following sets are definable: the set $\{0\}$, the set of all positive reals, and the set of all reals x such that $0 \leq x \leq 1$; these sets are determined respectively by the sentential functions '$\sigma(x, x, x)$', '$\neg\sigma(x, x, x).\exists y(\mu(y, x).\sigma(y, y, y))$', and '$\exists y \exists z(\mu(y, x).\sigma(y, y, y).\mu(x, z).\nu(z))$'.

So Tarski sketches here, for the first time, a rigorous definition of definability in metamathematical terms — in particular, in terms of the metamathematical notion of satisfaction. But now Tarski takes a further step: he offers a (partial) reconstruction of the notion of definability *in purely mathematical terms*.[44] Tarski writes:

> "Under this new definition the notion of definability does not differ from other mathematical notions and need not arouse either fears or doubts; it can be discussed entirely within the realm of normal mathematical reasoning." (p. 111)

This purely mathematical reconstruction rests on the following observation: each sentential function determines a certain set of finite sequences, namely, the set of all finite sequences that satisfy it. So we can replace the metamathematical concept of a sentential function by its mathematical analogue — the notion of a set of sequences.

In particular, we can introduce the *primitive* sets of sequences, those which are determined by the primitive sentential functions. Tarski goes on to define five operations on sets of sequences: complementation, addition, multiplication, and

[44] No total reconstruction is possible on pain of paradox; Tarski's partial reconstruction provides a mathematical definition of definable sets of individuals of order n, for a given n. The construction is carried out in detail for $n = 1$, and outlined for $n = 2$; this suffices, says Tarski, to make clear the method for all higher levels.

summation and multiplication with respect to the kth members of the sequences, corresponding to the five operations on sentential functions.[45] Then the family (or set) Df of definable sets of finite sequences of real numbers may be defined along the following lines:

> Df is the intersection of all the families of sets K which satisfy the following conditions: (i) the primitive sets belong to K; (ii) K is closed under the five operations.[46]

The family Df — characterized in purely mathematical terms — is exactly the family of sets of sequences which are determined by sentential functions. We have moved from metamathematical notions (sentential functions, satisfaction) to mathematical notions (sets of sequences, operations on sets of sequences).

Now we can move on to a mathematical definition of the notion *definable set of reals*. Among the sets of sequences that belong to Df there are all the sets U of unit sequences. For any such U, form the set U^* of the members of U's unit sequences. U^* is a *definable set of real numbers*. The set D of all these U^*s is the family of all *definable sets of real numbers*.[47] The set D is identical to the set of definable sets of real numbers as metamathematically defined by (M). Here, then, we have reached a purely mathematical characterization of the notion of a definable set of real numbers. The fact that it coincides with the metamathematical characterization demonstrates its material adequacy.

Tarski goes on to state two very important theorems about Df and D, and sketches their proofs.

Theorem. A set S of sequences of real numbers is a member of Df iff S is a finite sum of finite products of elementary linear sets.[48] (Theorem 1 in [Tarski, 1931a, p. 133].)

---

[45] The operations of complementation, addition, multiplication correspond to the sentential operations of negation, disjunction, and conjunction; they also correspond to the usual Boolean operations. The operations of summation and multiplication with respect to the kth terms correspond to the existential and the universal quantifiers respectively.

[46] See Definition 9 (p. 128) for a precise formulation.

[47] See Definition 10 (p. 128) for a rigorous formulation of the definition of definable sets of individuals.

[48] For the notion of an elementary linear set, we need the notion of a linear polynomial with integral coefficients, exemplified by $3x + 10y + 7z + 2$. Associated with this polynomial is a linear equation

(i) $3x + 10y + 7z + 2 = 0$ and a linear inequality

(ii) $3x + 10y + 7z + 2 > 0$.

Observe that associated with (i) are sequences of reals, each of which is a solution to (i). One such sequence is $\langle 2, -1.5, 1\rangle$, since $x = 2, y = -1.5$ and $z = 1$ provides a solution to (i). We can make a parallel observation about (ii).

An *elementary linear set* is a set of all sequences of real numbers which are solutions either of a linear equation of the type of (i), or a linear inequality of the type of (ii). (For a more rigorous characterization, see Tarski [Tarski, 1931a, p. 132].)

Theorem. A set X of real numbers belongs to the set D iff X is the sum of a finite number of intervals with rational end-points.[49] (Theorem 2 in [Tarski, 1931a, p. 133].)

Theorem 1 has a special significance. Given the correlation between (mathematical) sets of sequences and (metamathematical) sentential functions, and between Boolean operations on sets and Boolean operations on sentential functions, there is a metamathematical analogue of Theorem 1. And this metamathematical theorem leads to metalogical results of great moment: the theory of the reals $\langle R, 1, \leq, + \rangle$ is complete and decidable. More on this in Section 5.1 below.[50]

4.2 Truth

Tarski's seminal monograph 'The concept of truth in formalized languages' [Tarski, 1933a], and his later informal summary 'The Semantic Conception of Truth' [Tarski, 1944] are devoted to a single problem — the definition of truth. Tarski seeks a definition of truth that does justice to the classical Aristotelian conception of truth:

> "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true."[51]

Tarski writes:

> "Let us start with a concrete example. Consider the sentence "snow is white". We ask the question under what conditions this sentence is true or false. It seems clear that if we base ourselves on the classical conception of truth, we shall say that the sentence is true if snow is white, and that it is false if snow is not white. Thus, if the definition of truth is to conform to our conception, it must imply the following equivalence:
>
> The sentence "snow is white" is true if, and only if, snow is white." [1944, p. 667]

The account is readily generalized:

> "Let us consider an arbitrary sentence: we shall replace it by the letter 'p'. We form the name of this sentence and we replace it by another

[49] An *interval with rational end points* is a set of one of the following types, where a and b are arbitrary rational numbers: $\{x|x = a\}, \{x|x > a\}, \{x|x < a\}$, or $\{x|a < x < b\}$.

[50] Theorems 1 and 2 also have significance for analytic geometry. Near the end of [Tarski, 1931a, pp. 141–2], Tarski reports that Kuratowski drew his attention to a geometric interpretation of the concepts and operations of [Tarski, 1931a]. (For example, on this interpretation the operation of summation with respect to the kth members may be taken as projection parallel to the X_k axis.) Interpreted in this way, Theorems 1 and 2 take the form of certain theorems of analytic geometry. This geometric interpretation is investigated in [Tarski, 1931b].

[51] Aristotle, *Metaphysics*, Γ, 7, 27.

> letter, say 'X'. We ask now what is the logical relation between the two sentences "X is true" and 'p'. It is clear that from the point of view of our basic conception of truth these sentences are equivalent. In other words, the following equivalence holds:
>
> (T) X is true if, and only if, p. [Tarski, 1944, p. 668]

Tarski continues:

> "Now at last we are able to put into a precise form the conditions under which we will consider the usage and definition of the term "true" as adequate from the material point of view: we wish to use the term "true" in such a way that all equivalences of the form (T) can be asserted, and *we shall call a definition of truth "adequate" if all these equivalences follow from it*." [Tarski, 1944, p. 668]

Here, then, is an informal expression of Tarski's famous adequacy condition on a definition of truth.

However, Tarski stresses that

> "*the attempt to construct a correct semantical definition of the expression 'true sentence' meets with very real difficulties*." [Tarski, 1933a, p. 162]

This is primarily because of the well-known Liar paradox. Tarski presents the Liar along the following lines. Consider the following sentence:

> The sentence printed in this chapter on p. 542, line 21, is not true.

Let the letter 's' be an abbreviation of this sentence. Now, we may assert the following instance of schema (T):

> 's' is true iff the sentence printed in this chapter on p. 542, line 21, is not true.

We can establish empirically that

> 's' is identical to the sentence printed in this chapter on p. 542, line 21, is not true.

By the logic of identity, we obtain

> 's' is true iff 's' is not true.

And we arrive at a contradiction.

Tarski identifies two assumptions that are essential to the construction of the Liar. The first assumption is this:

> "(I) We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term "true" referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language." [Tarski, 1944, p. 672]

Tarski calls a language with these properties "semantically universal" [Tarski, 1969, p. 89],[52] or "semantically closed" [Tarski, 1944, p. 672]. The second assumption is this:

> "(II) We have assumed that in this language the ordinary laws of logic hold." [Tarski, 1944, p. 672]

A possible response is to reject II and give up classical logic. For Tarski, this is not an option:

> "It would be superfluous to stress here the consequences of rejecting the assumption II, that is, of changing our logic (supposing this were possible) even in its more elementary and fundamental parts. We thus consider only the possibility of rejecting the assumption I." [Tarski, 1944, p. 673]

And so Tarski concludes that we should seek a definition of truth for languages that are not semantically universal.

This leads to a distinction between *object language* and *metalanguage*:

> "The first of these languages is the language which is 'talked about' and which is the subject-matter of the whole discussion; the definition of truth which we are seeking applies to the sentences of this language. The second is the language in which we 'talk about' the first language, and in terms of which we wish, in particular, to construct the definition of truth for the first language." [Tarski, 1944, pp. 673–674]

Tarski points out that the terms 'object language' and 'metalanguage' apply only in a relative way:

> "If, for instance, we become interested in the notion of truth applying to sentences, not of our original object language, but of its metalanguage, the latter becomes automatically the object language of our discussion; and in order to define truth for this language, we have to go to a new metalanguage — so to speak, to a metalanguage of a higher level. In this way we arrive at a whole hierarchy of languages." [Tarski, 1944, p. 674]

[52] [Tarski, 1969] is a semi-popular account of the notions of truth and proof.

So in Tarski's view we can hope to construct a definition of truth for a language L only if L is not semantically universal, only if there is a metalanguage for L in which we can couch the definition. But now, according to Tarski, colloquial or natural languages (such as English or Polish) *are* semantically universal. In fact, according to Tarski, we can say something stronger: they are universal, in the sense that they can say everything that can be said. Tarski writes:

> "A characteristic feature of colloquial language (in contrast to various scientific languages) is its universality. It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that 'if we can speak meaningfully about anything at all, we can also speak about it in colloquial language'." [Tarski, 1933a, p. 164]

In the same vein, Tarski remarks on the "all-comprehensive, universal character" of natural language and continues

> "The common language is universal and is intended to be so. It is supposed to provide adequate facilities for expressing everything that can be expressed at all, in any language whatsoever; it is continually expanding to satisfy this requirement." [Tarski, 1969, p. 411]

It is the universality of natural languages, says Tarski, "which is the primary source of all semantical antinomies" [Tarski, 1933a, p. 164]. For if a language can say everything there is to say, then in particular, it can say everything there is to say about its own semantics. Such a language is semantically universal. But a semantically universal language, together with the usual laws of logic, leads to semantic paradox (see [Tarski, 1933a, pp. 164–5]). Tarski concludes:

> "The very possibility of a consistent use of the expression 'true sentence' which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the possibility of constructing a correct definition of this expression." ([Tarski, 1933a, p. 165]. See also p. 267)

So Tarski turns away from natural language.[53] He writes:

> "I now abandon the attempt to solve our problem for the language of everyday life and restrict myself henceforth entirely to *formalized languages*." [Tarski, 1933a, p. 165]

For the sake of clarity, Tarski constructs a definition of truth for one particular formal language — though as he points out, the method applies generally to an

[53] Despite Tarski's own misgivings, others have been less pessimistic — for example, Davidson in [Davidson, 1967], and a number of authors who offer a 'Tarskian' approach to the Liar in the sense that they attribute a hierarchical structure to natural languages (see, for example, Parsons [1974], Burge [1979], Barwise and Etchemendy [1987], Gaifman [1992], Koons [1992], Glanzberg [2001]).

extensive group of formalized languages. Tarski chooses as his object language the *calculus of classes*. This calculus is one interpretation of Boolean algebra, and a particularly simple fragment of mathematical logic. There are four constants — negation ('N'), disjunction or logical sum ('A'), the universal quantifier ('Π'), and inclusion ('I') — and an infinite supply of variables ('$x_{'}$', '$x_{''}$', '$x_{'''}$') which range over classes, and which may be arranged in a sequence (so that we may speak of the kth variable). Composite expressions, such as $Ix_{''}x_{'''}$, are formed in the obvious way.

The concept of truth for the calculus of classes is to be defined in a metalanguage. Accordingly,

> "the proper domain of the following considerations is not the language of the calculus of classes itself but the corresponding metalanguage. Our investigations belong to the *metacalculus of classes* developed in this metalanguage." [Tarski, 1933a, p. 169]

The metalanguage includes the logical expressions 'not', 'or', 'for all', and 'is included in', which have the same meanings as the constants of the object language — so the object language can be translated into the metalanguage. The metalanguage includes further expressions of a general logical character, including set-theoretical expressions (such as 'is an element of', 'class' and 'natural number') and expressions from the logic of relations (such as 'domain' and 'ordered n-tuple'). There is one further important category of expressions in the metalanguage: the expressions of a *structural-descriptive character*. These are the names of expressions of the object language. Tarski mentions the following names in particular: 'the negation sign' (abbreviated by 'ng'), 'the sign of logical sum' ('sm'), 'the sign of the universal quantifier' ('un'), 'the inclusion sign' ('in'), 'the kth variable' ('v_k'), and 'the expression which consists of two successive expressions x and y' ('$x^\frown y$'). Tarski observes that the metalanguage contains both an individual name and a translation of every expression (and so every sentence) of the object language; this is crucial for the construction of the definition of truth.

Tarski next turns to the metatheory, and presents the axiom system of the metalanguage. The axioms divide into two kinds. First there are the *general logical axioms* which provide a sufficiently comprehensive system of mathematical logic (we may take, for example, the axioms of *Principia Mathematica*). Axioms of the second kind deal with the structural-descriptive concepts. In his introduction to [83a], Corcoran stresses the significance of these axioms.

> "In the 'Wahrheitsbegriff', Tarski isolates as a primitive notion the fundamental operation of concatenation of strings and he presents, employing concatenation, the first axiomatic codification of string theory, thereby providing deductive foundations of scientific syntax." [Tarski, 1983a, p. xxi]

Tarski writes:

> "What we call metatheory is, fundamentally, the *morphology of language* — a science of the form of expressions — a correlate of such parts of traditional grammar as morphology, etymology, and syntax." [Tarski, 1933a, p. 251]

The axioms are as follows:

Axiom 1. ng, sm, un, and in are expressions, no two of which are identical.

Axiom 2. v_k is an expression if and only if k is a natural number distinct from 0; v_k is distinct from ng, sm, un, in, and also from v_l if $k \neq l$.

Axiom 3. $x^\frown y$ is an expression if and only if x and y are expressions; $x^\frown y$ is distinct from ng, sm, un, in, and from each of the expressions v_k.

Axiom 4. If x, y, z, and t are expressions, then we have $x^\frown y = z^\frown t$ if and only if one of the following conditions is satisfied: (a) $x = z$ and $y = t$; (b) there is an expression u such that $x = z^\frown u$ and $t = u^\frown y$; (c) there is an expression u such that $z = x^\frown u$ and $y = u^\frown t$.

Axiom 5. (The principle of induction) Let X be a class which satisfies the following conditions: (a) ng $\in X$, sm $\in X$, un $\in X$, and in $\in X$; (b) if k is a natural number distinct from 0, then $v_k \in X$; (c) if $x \in X$ and $y \in X$, then $x^\frown y \in X$. Then every expression belongs to the class X.

Axiom 5 tells us in a precise way that every expression consists of a finite number of signs.

In a series of definitions that follow these axioms (pp. 175–6), Tarski introduces three fundamental operations — negation ('$\bar{x}$' denotes the negation of x), disjunction ('$x + y$' denotes the disjunction of x and y), and universal quantification (denoted by '$\cap$'). Conjunction is defined in the usual way in terms of negation and disjunction, and the existential quantifier — denoted by '$\cup$' — is defined in terms of the universal quantifier and negation. By means of these operations, compound expressions are formed from simpler ones. This leads to a definition of the notion of a sentential function. The simplest sentential function is an *inclusion*, an expression of the form $(\text{in}^\frown v_k)^\frown v_l$, or $\iota_{k,l}$ for short. Compound sentential functions are formed by applying the three operations to inclusions any number of times. Sentences are then defined as a special case of sentential functions: they are sentential functions with no free variable. Corcoran remarks that Tarski here provides

> "the first formal presentation of a generative grammar" [Tarski, 1983a, p. xxi][54]

[54] Corcoran continues:

> 'It is to be regretted that many linguists, philosophers, and mathematicians know so little of the history of the methodology of deductive science that they attribute the basic ideas of generative grammar to linguists working in the 1950s rather than

Tarski now turns to the axioms for the object language, the calculus of classes. Among these are axioms that suffice for the sentential calculus, in which negation and disjunction are the only constants. The remaining axioms form a system for the calculus of classes, a simplified version of Huntington's system [Huntington, 1904, p. 297]. Tarski completes the preparations for his definition of truth by defining a number of basic metamathematical notions, including *consequence*, *theorem*, *deductive system*, *consistency*, and *completeness* (pp. 180–85).

The definition of *true sentence* in the language of the calculus of classes is carried out in Section 3 of [Tarski, 1933a]. At the outset of this section, Tarski returns to his adequacy condition and now states it in a more precise way. (In what follows, S is the class of all meaningful sentences of the language of the calculus of classes.)

> "*Convention T.* A formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called an *adequate definition of truth* if it has the following consequences:
>
> (α) all sentences which are obtained from the expression
>
> '$x \in$ Tr if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage;
>
> (β) the sentence 'for any x, if $x \in$ Tr then $x \in S$' (in other words 'Tr $\subseteq S$')." (pp. 187–8)

Tarski observes that if the object language contained only a finite number of sentences, it would be straightforward to produce an adequate definition. In schematic form, it would look like this:

$$x \in \text{Tr iff } (x = x_1 \text{ and } p_1) \text{ or } (x = x_2 \text{ and } p_2) \text{ or } \ldots \text{or } (x = x_n \text{ and } p_n)$$

where 'x_1', 'x_2', ... 'x_n' are replaced by structural-descriptive names of sentences of the object language, and 'p_1', 'p_2', ... 'p_n' by the corresponding translations of these sentences into the metalanguage. But if the object language contains infinitely many sentences, as is the case with the language of the calculus of classes, then we cannot proceed in this 'list-like' way — we cannot formulate infinitely long definitions in the metalanguage. Consequently, Tarski turns to the recursive method. Truth is to be defined in terms of the more basic notion of satisfaction, and satisfaction is defined recursively.

We saw above that in [Tarski, 1931a] Tarski sketches the notion of satisfaction. But now, for the first time in print, Tarski provides a rigorous definition of satisfaction (specific, of course, to the language of the calculus of classes).

> to Tarski (and other logician/methodologists) working in the early 1930s." [Tarski, 1983a, p. xxi]

Corcoran also remarks that although other authors had used ideas about scientific syntax prior to [Tarski, 1933a], Tarski was the first to present these ideas in a rigorous way — another instance where Tarski makes formally precise ideas already informally in use (see [Tarski, 1983a, p. xxi, fn. 13].

> "*Definition 22* The sequence f *satisfies* the sentential function x if and only if f is an infinite sequence of classes and x is a sentential function and if f and x are such that either (α) there exist natural numbers k and l such that $x = \iota_{k,l}$ and $f_k \subseteq f_l$; (β) there is a sentential function y such that $x = \bar{y}$ and f does not satisfy the function y; (γ) there are sentential functions y and z such that $x = y + z$ and f either satisfies y or satisfies z; or finally (δ) there is a natural number k and a sentential function y such that $x = \cap_k y$ and every infinite sequence of classes which differs from f in at most the kth place satisfies the function y." (p. 193)

It is remarkable how close this definition is to contemporary textbook definitions — that is a tribute to the precision and clarity of Tarski's work.

The definition of truth now follows. Members of sequences correspond to variables with respect to their indices. Whether or not a sequence satisfies a sentential function depends only on the members of the sequence that correspond to the free variables of the sentential function. In the limit case, where the sentential function has no free variables, the members of the sequence do not matter at all. As Tarski puts it:

> "Thus in the extreme case, when the function is a sentence, and so contains no free variable (which is in no way excluded by Def. 22), the satisfaction of a function by a sequence does not depend on the properties of the terms of the sequence at all. Only two possibilities then remain: either every infinite sequence of classes satisfies a given sentence, or no sequence satisfies it.... The sentences of the first kind ... are the *true sentences*; those of the second kind ... can correspondingly be called the *false sentences*." (p. 194)

So we arrive at Tarski's definition of truth:

> "*Definition 23* x is a true sentence — in symbols $x \in \mathrm{Tr}$ — if and only if $x \in S$ and every infinite sequence of classes satisfies x." (p. 195)

This definition is clearly formally correct — so the question now is whether it is materially adequate. The answer is in the affirmative: the definition is adequate in the sense of Convention T. Tarski points out that a rigorous proof of this would require a meta-metatheory, since it is a result about the metatheory. To avoid any such shift of levels, Tarski adopts an alternative method — the "empirical method", according to which we confirm the adequacy of the definition by way of concrete examples. The example Tarski considers is the sentence $\cap_1 \cup_2 \iota_{1,2}$. By successive applications of the relevant clauses of Definition 22, it is straightforward to derive:

> $\cap_1 \cup_2 \iota_{1,2} \in \mathrm{Tr}$ iff for every class a there is a class b such that $a \subseteq b$.[55]

[55] Following Tarski (p. 196), observe first that the sentential function $\iota_{1,2}$ is satisfied by exactly

This is a so-called 'T-sentence', one of the sentences mentioned in clause (α) of Convention T. And there is nothing special about this derivation — we can proceed in an analogous way with any sentence of the object language. Tarski concludes:

> "*We have succeeded in doing for the language of the calculus of classes what we tried in vain to do for colloquial language: namely, to construct a formally correct and materially adequate semantical definition of the expression 'true sentence'.*" (pp. 208–9, Tarski's emphasis)

Although Tarski has defined truth for a specific language, the method of construction applies to "many other formalized languages" (p. 209), as Tarski shows in Section 4. For each particular object language and object theory, we will need to construct a corresponding metalanguage and metatheory. The construction will follow the pattern indicated in the case of the calculus of classes. The metalanguage will always contain three groups of expressions: expressions of a general logical kind, expressions with the same meaning as the constants of the object language, and expressions of the structural-descriptive type. And the axiom system of the metatheory will include three groups of axioms: axioms of a general logical kind, axioms which provide for the translations into the metalanguage of all theorems of the object theory, and axioms for the structural-descriptive concepts. We will go on to define the notions of 'sentential function', 'sentence' and 'satisfaction', and define the notion of 'true sentence' in terms of satisfaction. Tarski concludes:

> "A. For every formalized language a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage with the help only of general logical expressions, of expressions of the language itself, and of terms from the morphology of language - but under the condition that the metalanguage possesses a higher order than the language which is the object of investigation." (postscript to [Tarski, 1933a, p. 273])

---

those sequences f such that $f_1 \subseteq f_2$. So the negation $\overline{\iota_{1,2}}$ is satisfied by exactly those sequences f such that $f_1 \not\subseteq f_2$. Consequently,

> "a sequence f satisfies the [sentential] function $\cap_2\overline{\iota_{1,2}}$ if every sequence g which differs from f in at most the 2nd place satisfies the function $\overline{\iota_{1,2}}$ and thus verifies the formula $g_1 \not\subseteq g_2$. Since $g_1 = f_1$ and the class g_2 may be quite arbitrary, only those sequences f satisfy the function $\cap_2\overline{\iota_{1,2}}$ which are such that $f_1 \not\subseteq b$ for any class b. If we proceed in an analogous way, we reach the result that the sequence f satisfies the function $\cup_2\iota_{1,2}$, i.e. the negation of the function $\cap_2\overline{\iota_{1,2}}$, only if there is a class b for which $f_1 \subseteq b$ holds. Moreover the sentence $\cap_1\cup_2\iota_{1,2}$ is only satisfied (by an arbitrary sequence f) if there is for an arbitrary class a, a class b for which $a \subseteq b$. Finally by applying Def. 23 we at once obtain one of the theorems which were described in the condition (α) of the convention T:
>
> $\cap_1 \cup_2 \iota_{1,2} \in$ Tr iff for every class a there is a class b such that $a \subseteq b$.
>
> From this we infer without difficulty, by using the known theorems of the calculus of classes, that $\cap_1 \cup_2 \iota_{1,2}$ is a true sentence." (p. 196)

> This condition on the metalanguage is crucial. For:
>
> "B. If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed." (p. 273)[56]

In particular, if the metalanguage is identical to, or interpretable in the object language, it is impossible to construct an adequate definition of truth. In Section 5, Tarski takes as object language the language of the general theory of classes, and considers a metalanguage which may be interpreted in the object language — so that every sentence of the metalanguage is correlated with an equivalent sentence of the object language. Suppose that under these circumstances we can construct in the metalanguage a correct definition of truth. Tarski writes:

> "It would then be possible to reconstruct the antinomy of the liar in the metalanguage, by forming in the language itself a sentence x such that the sentence of the metalanguage which is correlated with x asserts that x is not a true sentence." (p. 248)

Let us take a close look at Tarski's reconstruction of the liar in the metalanguage. We start with the language of the general theory of classes. The variables of this language range over individuals, classes of individuals, classes of classes of individuals, and so on. Following Tarski, let 'n' be a variable ranging over classes of classes of individuals, and treat numbers in the Russell–Whitehead way. Then the range of 'n' includes the natural numbers.[57] Tarski observes that we may construct a sentential function ι_k with sole free variable 'n' which says, given a natural number k, that the class whose name is represented by 'n' is identical with k (see p. 249).

Tarski further observes that it is straightforward to set up a correspondence between the expressions of the language and the natural numbers. That is, we can define in the metalanguage an infinite sequence $\varnothing$ of expressions in which every

[56] The order of a language is the smallest ordinal number which is greater than the orders of all variables occurring in the language (see p. 270). Variables ranging over individuals are assigned the order 0; variables ranging over classes of individuals and over relations between individuals are assigned the order 1; variables ranging over classes of classes of individuals and over relations between classes of individuals are assigned the order 2; and so on. The order of the language of the calculus of classes is 2, since its variables (ranging over classes of individuals) are of order 1. The language of the general theory of classes (introduced in Section 5) is a language of infinite order. Its variables range over classes, classes of classes, classes of classes of classes, and so on — and so its order is ω. (For the notion of the order of an expression, see the Postscript pp. 269–272, which revises Tarski's earlier account in section 4, p. 218).

Tarski's summary conclusion A from the Postscript is a significant revision of his summary in Section 6. In Sections 4 and 5, Tarski works with Husserl's notion of *semantical category*, introduced by Leśniewski into metamathematical investigations (Tarski cites Leśniewski [1929], especially pp. 14 and 68.) In the Postscript Tarski dispenses with the theory of semantical categories, and as a consequence finds that "the range of the results obtained has been essentially enlarged, and at the same time the conditions for their application have been made more precise" (p. 273).

[57] See Tarski p. 233, fn.1. For example, the number 1 is the class of all classes which have exactly one individual as a member.

expression occurs exactly once; we will let '$\varnothing_n$' denote the nth expression in the sequence. Consequently, we can correlate with every operation on expressions an operation on natural numbers, with every class of expressions a class of natural numbers, and so on. "In this way", Tarski writes,

> "the metalanguage receives an interpretation in the arithmetic of the natural numbers and indirectly in the language of the general theory of classes." (pp. 249–50)[58]

That is, the metalanguage is interpretable in the object language.

Now suppose that we have defined in the metalanguage the class Tr of true sentences of the object language. Let the symbol 'U^*y' denote the existential quantification of the sentential function y with respect to the variable 'n'. Consider the expression:

$$U^*(\iota_n.\varnothing_n) \notin \text{Tr}.$$

The expression '$\iota_n.\varnothing_n$' denotes the conjunction of the sentential functions denoted by ι_n and $\varnothing_n$. And '$U^*(\iota_n.\varnothing_n)$' denotes the existential quantification of this conjunction with respect to 'n'. As a whole, the expression '$U^*(\iota_n.\varnothing_{n)} \notin$ Tr' says that this existential quantification is not a true sentence of the object language.

The expression '$U^*(\iota_n.\varnothing_n) \notin$ Tr' itself is a sentential function of the metalanguage with 'n' as its sole free variable. Given that we can always correlate sentential functions with arithmetical functions, this sentential function may be correlated with another function equivalent to it but expressed solely in the terms of arithmetic. Let us schematically refer to this arithmetical function by '$\psi(n)$'. We have:

1. for any $n, U^*(\iota_n.\varnothing_n) \notin$ Tr if and only if $\psi(n)$.

Since the general theory of classes provides a foundation for the theory of natural numbers, it follows that '$\psi(n)$' is a sentential function of the object language. So it will appear in the sequence $\varnothing$ — let it appear at the kth place. Now instantiate (1) to 'k':

2. $U^*(\iota_k.\varnothing_k) \notin$ Tr if and only if $\psi(k)$.

Remember that we have assumed that the definition of Tr is adequate. Accordingly, condition (α) of Convention T holds. Now $U^*(\iota_k.\varnothing_k)$ is a structural-descriptive name of a sentence of the object language. The sentence that it names says the following: "there is an n such that $n = k$ and $\psi(n)$", or more simply, "$\psi(k)$". So by condition (α) we obtain:

3. $U^*(\iota_k.\varnothing_k) \in$ Tr if and only if $\psi(k)$.

[58] The theory of natural numbers can be developed in the general theory of classes. Tarski writes that the general theory of classes "suffices for the formulation of every idea which can be expressed in the whole language of mathematical logic" (pp. 241–2).

And (2) and (3) yield a contradiction.

Notice that '$\psi(k)$', a sentence of the object language, is correlated with '$U^*(\iota_k.\varnothing_k) \notin \mathrm{Tr}$', a sentence of the metalanguage, and the latter says of the former that it is not true. Thus, to repeat Tarski's words, we have formed in the object language itself "a sentence x such that the sentence of the metalanguage which is correlated with x asserts that x is not a true sentence." In reconstructing the Liar in this way, Tarski has used the so-called *diagonal argument*; the distinctive 'diagonal' step is the inference to (2), where we replace the variable 'n' in the correlates '$U^*(\iota_n.\varnothing_n) \notin \mathrm{Tr}$' and '$\psi(n)$' by k, the index of '$\psi(n)$' in $\varnothing$.[59] One can also see close analogies with the heterological paradox, generated by the phrase 'is not true of itself' or 'heterological' for short. Tarski pursues this analogy in [Tarski, 1944, p. 673, fn. 11].

We now have a sketch of the proof of Tarski's undefinability theorem for the specific case of the general theory of classes. Tarski states the result as follows:

> "*Theorem 1* (α) In whatever way the symbol 'Tr', denoting a class of expressions, is defined in the metatheory, it will be possible to derive from it the negation of one of the sentences which were described in the condition (α) of the convention T;
>
> (β) assuming that the class of all provable sentences of the metalanguage is consistent, it is impossible to construct an adequate definition of truth in the sense of convention T on the basis of the metatheory." [Tarski, 1933a, p. 247]

Clearly, (β) follows immediately from (α). For a general statement of the theorem, recall Tarski's summary conclusion in the Postscript:

> "B. If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed." (p. 273)

A typical contemporary formulation of Tarski's theorem is likely to dispense with any mention of a metalanguage; for example,

> Given a language L of a theory that includes first-order number theory, no adequate definition of truth in L can be given in L.

Of his proof of Theorem 1, Tarski writes "We owe the method used here to Gödel" (fn. 1, p. 247). Indeed, (α) does not really involve truth at all.[60] If we replace the symbol 'Tr' by the more neutral symbol 'ϕ', (α) tells us that there is a sentence p with a structural-descriptive name x such that the sentence '$\neg(\phi(x) \leftrightarrow p)$' and hence '$\phi(x) \leftrightarrow \neg p$' is derivable from the metatheory. If we interpret ϕ not as 'true' but as 'provable', (α) is essentially an instance of Gödel's first incompleteness theorem. Originally, Tarski's paper did not contain Theorem 1,

[59] For more on the diagonal argument, the Liar, and Tarski's theorem, see [Simmons, 1993, Chapter 2].

[60] This point is stressed by Vaught in [Tarski, 1986a, p. 871].

but instead "certain suppositions in the same direction" (fn. 1, p. 247), based partly on Tarski's own work and partly on a short report of Gödel's [Tarski, 1930b]. Theorem 1 and the sketch of its proof were added in press, after Tarski became acquainted with [Gödel, 1931] and was convinced that Gödel's results carried over to the general theory of classes.[61]

In notes added to [Tarski, 1933a], Tarski is at pains to set the historical record straight. There was some need for this. For example, Carnap erroneously wrote that Tarski's investigations were carried out "in connection with those of Gödel" ([Carnap, 1935]). In fact Tarski arrived at the final formulation of his definition of truth in 1929. But the paper was not published until 1933 (in Polish) — and only received more general attention when it was published in German in 1935. Tarski emphasizes that the methods and results of [Tarski, 1933a] were entirely his own, with two noted exceptions — Section 2 (on the problems of defining truth in natural language) was directly influenced by Leśniewski, and the method of the proof of Theorem 1 was drawn from Gödel [1931]. It is noteworthy that Tarski's discussion of the arithmetization of the metalanguage and metatheory proceeded quite independently of Gödel's work (though Gödel developed the method far more completely).

****

Tarski's theory of truth has received far more philosophical attention than anything else in Tarski's corpus. It is without doubt the work for which Tarski is best known to philosophers. It is also the subject of Tarski's most sustained philosophical discussion, which comprises the second half of [Tarski, 1944] (under the title of 'Polemical Remarks'). We turn now to some of the philosophical issues raised by Tarski's work on truth.

(i) There is disagreement over the *point* of Tarski's definition of truth. What is Tarski's aim? To this question, I believe there is a clear answer in Tarski's writings. On the one hand, truth is a notion that is fundamental to science, logic and metamathematics; on the other hand, it leads to paradoxes and antinomies, earning itself an "evil reputation". We must then find a way of defining truth that will preserve its intuitive content but place it beyond suspicion. Tarski seeks a definition of truth in terms that no one could question. And for Tarski, those terms are the terms of logic, the terms of the object language, and the structural-descriptive terms:

> "... the semantical concepts are ... reduced to purely logical concepts, the concepts of the language being investigated and the specific concepts of the morphology of language" [Tarski, 1936a, p. 406]

Moreover, this reduction is strongly suggested by the T-sentences. According to Tarski, the T-sentence

[61] These negative results about truth and provability are informally discussed in Tarski [Tarski, 1969].

> 'snow is white' is true if and only if snow is white

is a partial definition of truth — and it contains a logical term, a sentence of the object language, and a structural-descriptive name of that sentence. As we saw earlier, Tarski suggests that for an object language with finitely many sentences, we need only list all the associated T-sentences for a complete definition of truth. This will not work, of course, for languages with infinitely many sentences — for those, Tarski uses the recursive method. But still, says Tarski, his recursive definition effects the kind of reduction suggested by the T-sentences (see for example [Tarski, 1933a, pp. 251–3]).

Suppose now that in the metalanguage we introduce semantic concepts (referring to the object language) by way of Tarski's definition. Then:

> "the definition of truth, or of any other semantic concept, will fulfil what we intuitively expect from every definition; that is, it will explain the meaning of the term being defined in terms whose meaning appears to be completely clear and unequivocal. And, moreover, we have then a kind of guarantee that the use of semantic concepts will not involve us in any contradictions." [Tarski, 1944, p. 675]

Here is the point of Tarski's definition. In the metalanguage, we may use the notion of truth (for the object language) without fear of paradox, safe in the knowledge that 'true' can always be eliminated via its *definiens* in a precise, uncontroversial way, in favour of terms which are not semantical and thereby immune to paradox.[62]

Tarski's definition, then, legitimates the use of the concept of truth (*for* a given object language, *in* a suitable metalanguage). And Tarski was optimistic that the concept of truth thus defined would prove fruitful, both for philosophy and the special sciences. Tarski points out that the problem of defining truth "has often been emphasized as one of the fundamental problems of the theory of knowledge" [Tarski, 1936a, p. 407]. He suggests that the notion of truth is essential to the following constraint on empirical theories:

[62]These remarks are in broad agreement with Etchemendy's discussion in section 1 of [Etchemendy, 1988a]. They are opposed to Field's claim in [Field, 1972] that Tarski aimed for a physicalistic reduction of truth. Field rests his textual case on one brief passage from [Tarski, 1936a]:

> "... it would then be difficult to bring this method into harmony with the postulates of the unity of science and of physicalism (since the concepts of semantics would be neither logical nor physical concepts)." [Tarski, 1936a, p. 406]

The method in question is the axiomatic method of defining truth, according to which truth is introduced as a primitive concept whose basic properties are established by axioms. Tarski's concern here is that a primitive semantic concept would be in tension with the scientific, physicalist outlook. Tarski's theory produces no such tension: truth is eliminable, not primitive. Immediately after the passage just quoted, Tarski remarks that his preferred method reduces the concept of truth to logical concepts, concepts of the object language, and structural-descriptive concepts. There is no mention of physicalistic concepts or the need for a specifically physicalistic reduction. Tarski thought it an advantage of his theory that it harmonized with the scientific outlook — but there is no indication that Tarski sought an explicitly physicalist reduction of truth.

> "As soon as we succeed in showing that an empirical theory contains (or implies) false sentences, it cannot be any longer considered acceptable." [Tarski, 1944, p. 691]

And in regard to mathematics and metamathematics, Tarski suggests that we can already see the valuable results contributed by the theory of truth and the semantic method:

> "These results concern the mutual relations between the notion of truth and that of provability; establish new properties of the latter notion (which, as well known, is one of the basic notions of metamathematics); and throw some light on the fundamental problems of consistency and completeness. ...
>
> Furthermore, by applying the method of semantics we can adequately define several important metamathematical notions which have been used so far only in an intuitive way — such as, e.g., the notion of definability or that of a model of an axiom system; and thus we can undertake a systematic study of these notions. In particular, the investigations on definability have already brought some interesting results, and promise even more in the future." (pp. 692–693)

(ii) What kind of theory is Tarski's theory of truth? Is it, for example, a *correspondence* theory of truth? At several places, Tarski says that he wishes to capture the intuitions behind the classical Aristotelian conception of truth — and Tarski formulates that conception in terms of correspondence. Consider:

> "If we wished to adapt ourselves to modern philosophical teminology, we could perhaps express this conception by means of the following formula:
>
> > The truth of a sentence consists in its agreement with (or correspondence to) reality." ([Tarski, 1944, p. 667]. See also [Tarski, 1933a, p. 153] and [Tarski, 1936a, p. 401].)

However, we would expect a full-blown correspondence theory to provide an account of the correspondence relation and its relata — and Tarski's theory does not. Such notions as *correspondence* and *state of affairs* are not the subject of Tarski's theory — for Tarski, these are vague notions that gesture towards a certain conception of truth. It is noteworthy that Aristotle's dictum (quoted above) makes no mention of a correspondence relation, or facts, or states of affairs. And Tarski finds Aristotle's formulation preferable to those that talk vaguely of 'correspondence' and 'reality' (see [Tarski, 1944, p. 667]). Still, Aristotle's formulation, like the others, cannot serve as a definition of truth, and

> "[i]t is up to us to look for a more precise expression of our intuitions." [Tarski, 1944, p. 667]

According to Tarski, those intuitions are adequately accommodated by the T-sentences. Take the sentence 'snow is white'. To what fact or state of affairs or condition does this sentence correspond? To the fact, state of affairs or condition that snow is white. That is, the sentence is true if snow is white, and false if snow is not white — or to put it in familiar terms, 'snow is white' is true if and only if snow is white. A definition of truth conforms to the correspondence conception if it implies such T-sentences. (See [Tarski, 1944, p. 667]). So we can attribute to Tarski a correspondence *conception* of truth — but we should not say that Tarski's *theory* is a correspondence theory in any robust sense.

If Tarski's theory is not a correspondence theory, is it a kind of *deflationary* theory?[63] Deflationists, in particular disquotationalists, have drawn heavily from Tarski's account. According to the disquotationalist (for example Quine, in [Quine, 1970, Chapter 1], saying

> 'snow is white' is true

is just an indirect way of saying something about the world, namely, that snow is white. All there is to say about the truth of 'snow is white' is encapsulated by its associated T-sentence. According to the disquotationalist, truth is not a substantive property; rather, the term 'true' is a "device for disquotation" ([Quine, 1970, p. 12]). A natural disquotational definition of 'true' for a language L with finitely many sentences is this:

> x is true iff ($x = x_1$ and p_1) or ($x = x_2$ and p_2) or ... or ($x = x_n$ and p_n)

where 'x_1', 'x_2', ... 'x_n' are replaced by quote-names of sentences of L, and 'p_1', 'p_2', ... 'p_n' by the sentences themselves. And this definition looks very like the one Tarski suggests for languages with finitely many sentences (see above). So there appear to be some close connections between Tarski's theory and disquotationalism.

Of course, Tarski and disquotationalists are concerned with languages that contain infinitely many sentences. For that reason, as we saw, Tarski employs the recursive method. The disquotationalist might follow suit: given a language with a finite stock of names and predicates, reference may be disquotationally defined by a finite list of sentences of the form ' "a" refers to a', and satisfaction by a finite list of sentences of the form 'x satisfies "F" if and only if x is F'. But such a recursive disquotationalism is restricted to sentences that have the appropriate kind of logical form. And there is an array of truths that are notoriously hard to fit into the Tarskian mould: belief attributions, counterfactuals, modal assertions, statements of probability, and so on. Since disquotationalism is an account of the truth predicate of an entire natural language and not of some restricted portion, the recursive route seems unattractive.

[63] For a recent collection of classic and contemporary papers on deflationism see [Blackburn and Simmons, 1999].

The disquotationalist might prefer to extend the above finitary definition to an infinitary one:

x is true iff ($x = x_1$ and p_1) or ($x = x_2$ and p_2) or ...

where 'p_1', 'p_2' ... abbreviate the sentences of the language.[64] So now 'true' is not just a device of disquotation; it is also a device for expressing infinite disjunctions. Here the disquotationalist and Tarski part company; Tarski writes:

> "Whenever a language contains infinitely many sentences, the definitions constructed automatically according to the above scheme would have to consist of infinitely many words, and such sentences cannot be formulated either in the metalanguage or in any other language." [Tarski, 1933a, p. 188–9][65]

But Tarski's infinitary qualms aside, might it be claimed that disquotationalism captures the deflationary spirit of Tarski's theory?

Such a claim might be further encouraged by the observation that Tarski's definition is eliminative. According to the disquotationalist we have just characterized, 'true' is always in principle eliminable by disquotation. The *definiens* contains, apart from logical terms, only the quote-names of the sentences of the given language, and their disquotations. And Tarski is quite clear that his definition of truth is eliminative (see [Tarski, 1944, p. 683]). Tarski points out that the elimination may not be as straightforward as it is with " 'snow is white' is true" — consider the sentences "All consequences of true sentences are true" and "The first sentence written by Plato is true", where the elimination of 'true' is not a matter of simple disquotation (see [Tarski, 1944, p. 683]). Nevertheless, elimination is always possible in principle:

> "Of course, since we have a definition for truth and since every definition enables us to replace the definition by its *definiens*, an elimination of the term 'true' in its semantic sense is always theoretically possible." [Tarski, 1944, p. 683]

However, despite the close connections between Tarski and the disquotationalist, and the shared eliminative nature of their definitions, I think Tarski would strongly resist the label 'deflationist'. First, unlike the disquotationalist, Tarski is not offering a general theory of truth in natural languages. His theory is a limited to certain regimented languages — including fragments of a natural language, but never the whole language. Second, for Tarski truth is a substantive and fruitful

[64] Such a definition is suggested by remarks in [Leeds, 1978], and versions of it are presented explicitly in [Field, 1986], [Resnik, 1990], [David, 1994].

For an illustration, consider the sentence: 'What Claire said yesterday was true'. According to the definition, the sentence is equivalent to What Claire said='aardvarks amble' and aardvarks amble or What Claire said = 'antelopes graze' and antelopes graze or ....

[65] In later work, Tarski investigated infinitary languages — see [Tarski, 1958a; Tarski, 1958b; Tarski, 1961b; Tarski, 1961c].

concept. According to the disquotationalist, there is no substantive concept of truth — only the term 'true' that serves as a logical device. But, according to Tarski, as we have seen, truth is a concept that has important work to do — in philosophy and in metamathematics, for example. Third, Tarski thinks it is a mistake to think that the concept of truth is sterile on the grounds that the word 'true' may be eliminated on the basis of its definition. That would lead, Tarski says, to the absurd conclusion that all defined notions are sterile. Tarski writes:

> "In fact, I am rather inclined to agree with those who maintain that the moments of greatest creative achievement in science frequently coincide with the introduction of new notions by means of definition." [Tarski, 1944, p. 683]

Tarski reserves the label 'the semantic conception of truth' for his account. And the label is useful, at least for distinguishing Tarski's theory from correspondence and deflationary theories. Tarski's theory does not fit comfortably into either of those categories. Tarski himself emphasizes the neutrality of his theory:

> "we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians — whatever we were before. The semantic conception is completely neutral toward all these issues" [Tarski, 1944, p. 686]

Epistemological and metaphysical disputes about truth are not Tarski's concern. Rather, Tarski is after a definition of 'true sentence' that is precise, fruitful, immune to paradox, and consonant with ordinary usage.

(iii) There are many other philosophical issues raised by Tarski's work on truth. Here we indicate just some of them.

(a) Is truth immanent or transcendent? It is striking that Tarski defines 'true' only relative to a given language:

> "We can only meaningfully say of an expression that it is true or not if we treat this expression as a part of a concrete language. As soon as the discussion concerns more than one language the expression 'true sentence' ceases to be unambiguous." [Tarski, 1933a, p. 263]

The question arises whether truth must always be 'immanent' in this way, or whether there is a more general notion that transcends particular languages.[66]

(b) Are natural languages universal in the way that Tarski says they are? At issue here is the expressive capacity of natural languages.[67]

(c) Are natural languages inconsistent? Tarski is often taken to give an affirmative

[66] For discussions of this issue, see for example [Resnik, 1990] and [Field, 1994].

[67] For further discussion see, for example, [Martin, 1976], [Herzberger, 1981] and [Simmons, 1993].

answer to this question, and he certainly comes close. He says that the semantical antinomies

> "seem to provide a proof that every language which is universal in the above sense, and for which the normal laws of logic hold, must be inconsistent." [Tarski, 1933a, pp. 164–5]

But note the "seem to". In [Tarski, 1944], Tarski is more explicitly cautious — since a natural language has no exactly specified structure, "the problem of consistency has no exact meaning with respect to this language" [Tarski, 1944, p. 673]. Tarski also notes in [Tarski, 1933a, p. 267] that bringing exact methods to bear on natural language would require a reform of the language, including its division into a sequence of object languages and metalanguages.

> "It may, however, be doubted whether the language of everyday life, after being 'rationalized' in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalized languages." [Tarski, 1933a, p. 267]

There is, then, a delicate question as to how to read Tarski here. My preferred reading is this: it is inappropriate to bring exact methods to bear on natural language, and if one insists on doing so, the result will either be confusions and contradictions or an artificial regimentation of natural language.[68]

(d) How is Tarski's definition of truth related to the truth-conditions theory of meaning? According to [Davidson, 1967], the relationship is very close:

> "There is no need to suppress, of course, the obvious connection between a definition of truth of the kind Tarski has shown how to construct, and the concept of meaning. It is this: the definition works by giving necessary and sufficient conditions of the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence. To know the semantic concept of truth for a language is to know what it is for a sentence — any sentence — to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language." (p. 76)

Others would deny the connection that Davidson maintains here.[69] It is clear that Tarski would not regard his definition of truth as the basis for a theory of meaning for natural languages. For one thing, Tarski restricts himself to formal languages. And for another, truth, not meaning, is Tarski's target. The meanings of the sentences of the object language and metalanguage are taken as given from the outset — indeed, it is a requirement of Tarski's definition that the object language be translatable into the metalanguage.

[68] For further discussion see, for example, [Soames, 1999, pp. 49–56 and 150–51].

[69] See for example Etchemendy [1988a, pp. 56–62], and Soames [1999, pp. 102–107]. Davidson's own position on this issue evolved over the years — for example, one may compare and contrast Davidson [1967] and Davidson [1996].

4.3 Logical consequence

We have just seen how Tarski took up the challenge of providing a precise definition of truth that conforms to ordinary usage. In his paper "On the concept of logical consequence" [Tarski, 1936b], Tarski takes up the parallel challenge for the concept of *logical consequence*. At the outset of the paper, Tarski rejects the idea that logical consequence can be captured exhaustively via formal rules of inference. The idea had great appeal among logicians: after all, all exact reasonings in mathematics were formalizable using a small stock of simple rules of inference. But Tarski showed that there is a clear and unbridgeable gap between the ordinary notion of logical consequence and the logicians' formalized notion.

As early as 1927, Tarski showed the existence of so-called *ω-incomplete theories* (the terminology, but not the concept, is due to Gödel [Gödel, 1931]).[70] Such a theory includes the normal rules of inference (such as substitution and detachment), and among its theorems are the following:

A_0. 0 possesses the property P

A_1. 1 possesses the property P

and in general every sentence of the form

A_n. n possesses the property P.

But the following sentence cannot be proved in the theory:

A. Every natural number possesses the property P.

Now it is clear that, according to our ordinary notion, A *is* a logical consequence of $A_0, A_1, \ldots A_n, \ldots$. So:

> "the formalized concept of consequence, as it is generally used by mathematical logicians, by no means coincides with the common concept." (p. 411)

Tarski goes on to point out that this gap cannot be closed by the introduction of new (infinitary) rules of inference. For Gödel's results show that

> "[i]n every deductive theory (apart from theories of a particularly elementary nature), however much we supplement the ordinary rules of inference by new purely structural rules, it is possible to construct sentences which follow, in the usual sense, from the theorems of this theory, but which nevertheless cannot be proved in this theory on the basis of the accepted rules of inference." [Tarski, 1936b, p. 412–413]

[70] Tarski reports (in [Tarski, 1933b, p. 279, fn. 2] that in 1927 he had delivered a paper entitled 'Remarks on some notions of the methodology of the deductive sciences' to the Second Conference of the Polish Philosophical Society in Warsaw in which he stressed the importance of the concepts of ω-consistency and ω-completeness, and communicated an example of an ω-incomplete theory. A slightly altered form of this theory appears in [Tarski, 1933b].

Consequently,

> "In order to obtain the proper concept of consequence, which is close in essentials to the common concept, we must resort to quite different methods and apply quite different conceptual apparatus in defining it." (p. 413)

The new methods are Tarski's semantic methods. While Tarski makes "no very high claim to complete originality" [Tarski, 1935, p. 414], it is clear that he regards his account as the first fully precise semantic treatment of logical consequence. Tarski's positive account in [Tarski, 1936b] starts out with two guiding intuitions. Suppose we have a class K of sentences and a sentence X which follows from K. The intuitions are these:

(i) It can never happen that every member of K is true and X is false.

(ii) The subject matter of K and X (the objects to which reference is made) have no effect on the consequence relation — only the *form* of the sentences matters.

(i) and (ii) suggest the following condition on the consequence relation:

(F) Given K and X, if the nonlogical constants are replaced uniformly, yielding K' and X', then X' must be true if all the sentences of K' are true.

Now Tarski points out that while (F) can serve as a necessary condition, it is not sufficient. The difficulty is this: we may not assume that the language contains designations for all possible objects.[71] (Here, Tarski's remarks are aimed at Carnap's definition of logical consequence.)[72] To avoid this difficulty, Tarski turns to the notions of sentential function and satisfaction. Assume a correspondence between variables and extralogical constants. Given a class L of sentences, replace each occurrence of an extralogical constant by its corresponding variable. We will obtain a class of sentential functions — call it L'. Given the concept of the satisfaction of a sentential function by a sequence of objects, we can define the notion of a *model* of L as follows:

> An arbitrary sequence S of objects is a *model* of the class L of sentences if and only if S satisfies every member of L'.

[71] Tarski points out that it may, and does, happen that X does not follow in the ordinary sense from the sentences of K, and yet the condition F is satisfied — where F is satisfied because the language with which we are dealing does not have a sufficient stock of extra-logical symbols. Tarski writes that the condition is sufficient

> "only if the designations of all possible objects occurred in the language in question. This assumption, however, is fictitious and can never be realized" [Tarski, 1936b, p. 416]

[72] In [Carnap, 1934, pp. 137ff], Carnap provided definitions of logical consequence and further derivative concepts. Tarski regarded them as inadequate because the defined concepts are dependent on the richness of the language under investigation.

The notion of a *model of a sentence* is to be understood in the the obvious way. And now Tarski states the definition of logical consequence:

> "The sentence X *follows logically* from the sentences of the class K if and only if every model of the class K is also a model of the sentence X." [Tarski, 1936b, p. 417]

This definition, says Tarski, agrees with common usage, and meets the conditions (i) and (ii) — for it follows from the definition that every consequence of true sentences is true, and that the consequence relation is independent of the meanings of extra-logical constants. Notice that a definition of *logical truth* is immediately forthcoming: a sentence Y is logically true if every sequence of objects is a model of Y.

There has been far less philosophical discussion of Tarski's notion of logical consequence than of his notion of truth. (Perhaps, as Gómez-Torrente suggests in [Gómez-Torrente, 1996, p. 126], this is a sign of its success — it is generally regarded as a standard fixture of modern logic.) But there has been some discussion, most notably the critical work of Etchemendy in [Etchemendy, 1983; 1988a; 1988b; 1990]. Etchemendy challenges the model-theoretic treatment of logical consequence, a treatment that is now quite standard.[73] Etchemendy also challenges the widespread attribution of this standard treatment to Tarski (see [Etchemendy, 1988a]). He argues that Tarski's definition of logical consequence diverges from the standard one.

According to Etchemendy, there are two main points of divergence. The first concerns the ω-rule of inference, the inference that allows us to move from $A_1, A_2, \ldots A_n, \ldots$ to A. Given a first-order formalization of arithmetic (where the numerals are nonlogical constants) the ω-rule is invalid according to the standard model-theoretic account — for there are interpretations of 'natural number' and '0', '1', '2', ... for which $A_1, A_2, \ldots A_n, \ldots$ are all true and A is false. But Tarski claims that the inference is intuitively valid, and that his definition respects this intuition. If so, then it follows that Tarski's definition and the standard model theoretic definitions are different. And Etchemendy goes on to argue that the price Tarski pays for the validation of the ω-rule is a trivialization of his analysis.[74]

The second main point of divergence concerns domain variation: according to Etchemendy, Tarski's account of logical consequence, unlike the standard account, does not require that we vary the domain of quantification. But now Tarski's account gets into trouble, Etchemendy argues. Suppose the intended interpretation

[73] For a defence of the model-theoretic account against Etchemendy's attack, see [Soames, 1999, pp. 117–136].

[74] According to Etchemendy, Tarski allows a nonstandard choice of logical constants, which may include the numerals and the quantifier 'every natural number'. Then A will be a consequence of $A_1, A_2, \ldots A_n, \ldots$, since "any set that contains each natural number contains every natural number" [Etchemendy, 1988a, p. 73]. Etchemendy argues that this flexibility with regard to the choice of logical constants leads to a certain trivialization of Tarski's analysis. However, below we will consider another reading of Tarski, according to which Tarski's definition does not diverge from the standard model-theoretic one, Tarski does not endorse nonstandard choices of logical constants, and no such trivialization results.

of a given language has a domain with two or more individuals. Now consider the sentence $\exists x \exists y(x \neq y)$. If we do not require domain variation, this sentence will be a logical truth.

If Etchemendy is right, Tarski's definition — to its detriment — diverges from the standard model-theoretic account in two important respects. But it is possible to resist Etchemendy's reading of Tarski. According to Gómez-Torrente [Gómez-Torrente, 1996], the supposed divergences are illusory. First, Tarski did not think that first-order versions of the ω-rule are valid — only certain versions in broader logical frameworks are valid. (Of course, an argument may be valid under some but not all of its formalizations). Gómez-Torrente argues that in [Tarski, 1936b] Tarski has in mind a presentation of arithmetic in a broad logical framework such as the simple theory of types or the calculus of classes. (We saw above in fn. 71 that Tarski discusses ω-incompleteness in [Tarski, 1933b], where the logical framework is the general theory of classes.) In such a broad framework, arithmetical expressions (like the numerals) are defined expressions, not primitives or nonlogical constants. Now consider the ω-rule of inference again. If there are any nonlogical constants, they will appear in P only — nowhere else are there any constants subject to reinterpretation. And since the class of logically defined natural numbers will coincide with the extension of the predicate 'natural number', A will follow from $A_1, A_2, \ldots A_n, \ldots$ (to borrow Etchemendy's words, quoted in fn. 74, any set that contains each natural number contains every natural number).

Second, although it is true that [Tarski, 1936b] makes no mention of domain variation, Gómez-Torrente provides several reasons for thinking that it is implicit in Tarski's definition. Perhaps the most compelling reason is the following. Tarski often specifies a domain (or universe of discourse) by way of a nonlogical predicate. For example, recall the 'textbook' introduction of a model in [Tarski, 1941a, Chapter VI] — detailed in 2.5 above. We start with a simple deductive theory, the theory of congruence. The nonlogical predicate 'S' denotes the set of all line segments — the domain or universe of discourse. Now the two axioms are *relativized to* S; for example, Axiom 1 says that for all $x \in S$, $x \cong x$. Now we abstract away from our particular theory. In particular, we replace 'S' by 'K', so that now the axioms are fully abstract, and relativized to 'K'. Now 'K' may be interpreted as the set of all line segments — but it may also be interpreted as the universal class, or the set of all numbers, or any set of numbers, and so on. In each case, the interpretation of 'K' will be the domain of the model. Here, and elsewhere in Tarski's work,[75] domain variation is accommodated by the reinterpretation of

[75]There are many other places where Tarski specifies a domain via a nonlogical predicate. For another example, consider Tarski's presentation of abstract Boolean algebra in [Tarski, 1935+1936] — detailed in 2.4 above. As we saw, the meaning of the predicate 'B' is given by 'the universe of discourse'. In the course of [Tarski, 1935+1936], 'B' is variously interpreted as the universe of sets (Boolean set algebra), the set S of all sentences (the calculus of sentences), and the class D of deductive systems (the calculus of systems). Again, domain variation is accommodated by the reinterpretation of a nonlogical predicate.

As further examples, Gómez-Torrente also cites all the first order theories, and the second-order theory of real arithmetic, presented in [Tarski, 1936c].

nonlogical predicates. But Tarski's test for logical consequence just is a matter of interpretation and reinterpretation — and so Tarski's definition as it stands accommodates domain variation.[76]

We have good reason, then, to maintain the received view: Tarski provided us with the first precise model-theoretic treatment of logical consequence — and Tarski's treatment remains standard to this day.

4.4 Model theory: some historical remarks

Tarski's seminal work on definability, truth and logical consequence were of central importance to the development of *model theory* — the study of the relation between formal languages and their interpretations. Chang and Keisler point out that model theory is a young subject, "not clearly visible as a separate area of research until the early 1950s" [Chang and Keisler, 1973, p. 3]. [Löwenheim, 1915] is generally regarded as the first specific contribution to model theory. In this paper, Löwenheim introduces *the method of elimination of quantifiers* (more on this below). Löwenheim also proves the theorem that bears his name: If a first-order sentence A is true in every finite domain but not every domain, then there is a denumerable domain in which A is not true. Löwenheim's result was improved by Skolem in [Skolem, 1920], yielding the Löwenheim-Skolem theorem: Any set of first-order sentences which has a model has a denumerable model. In [Skolem, 1922], Skolem went on to improve the result further: Any model of a set of first-order sentences has a denumerable submodel.[77]

Tarski's earliest contributions to model theory were presented in the university lectures and seminars that he gave at Warsaw University in the years 1926–1928 (see [Tarski, 1948a, p. 50, fn. 11]). It is reported that in 1927-8 Tarski produced further improvements to the Löwenheim-Skolem theorem as follows:

(i) Any denumerable set of first-order sentences that has an infinite model has an uncountable model.[78]

(ii) Any denumerable set of first-order sentences that has an infinite model has a

[76] Gómez-Torrente mentions two other reasons for thinking that domain-variation is implicit in Tarski's definition. First, as both Etchemendy and Gómez-Torrente point out, Tarski had already defined the notion of 'true sentence in an individual domain a' in [Tarski, 1933a], showing an explicit concern with domain variation prior to [Tarski, 1936b]. Why unnecessarily attribute to Tarski an abrupt change of mind? Second, [Tarski, 1936b] is a summary of an address given to a philosophical audience (it was delivered at the International Congress of Scientific Philosophy, in Paris, 1935). The paper is indeed informally written and brief — so it would not be surprising if Tarski suppressed certain more technical matters. (Gómez-Torrente [1996] is a careful and convincing paper which contains more argumentation and historical information that I am able to cover here.)

[77] Given a model M with domain D, a submodel M' of M has for its domain a subset of D. M' assigns to the expressions of the given first-order language restrictions of the extensions that M assigns — restrictions tailored to M''s smaller domain.

[78] According to the editors of Fundamenta Mathematicae, Tarski proved this result in 1927–8. See the editors' Note after the paper Skolem [1934].

model in each infinite power.[79]

Vaught reports that Tarski never published a proof of (i) or (ii); Tarski's proofs remains a mystery.[80] Tarski also explored the method of eliminating quantifiers, a powerful method for establishing metamathematical properties of theories, especially decidability. We will take a closer look at the method in Section 5 below. A number of key notions of model theory were introduced in these seminars — among them, Vaught reports, the notion of *elementary equivalence*.[81]

Gödel's landmark completeness theorem was published in 1930. During the 1930s, Tarski produced his greatest contributions to model theory — the work on definability, truth and logical consequence. As we have seen, Tarski stresses semantical methods in Chapter VI of his 1941 textbook [Tarski, 1941a], and the notion of model takes center-stage.

It was not until the postwar period that model theory became an autonomous area of study. In 1954–5, Tarski published a series of three articles under the title "Contributions to the theory of models" [Tarski, 1954a; 1954b; 1955]. In the first of these, Tarski writes:

> "Within the last few years a new branch of metamathematics has been developing. It is called the *theory of models* and can be regarded as a part of the semantics of formalized theories." [Tarski, 1954a, p. 517]

This appears to be the first time that the phrase 'theory of models' appears in the literature.[82] The subject developed quickly in the 1950s, stimulated by work of Tarski, Mal'cev, Henkin and Abraham Robinson, among others. Tarski continued to make important contributions: for example, he introduced the notions of *elementary substructures* and *elementary chains*, and proved some fundamental results involving these notions, including "Tarski's union theorem" (see [Tarski, 1957b], with Vaught); with Frayne, Morel and Scott, he showed the importance of the role of ultraproducts in model theory (see [Tarski, 1958c]); and with Hanf, he initiated work on measurable cardinals ([Tarski, 1962] and [Hanf, 1963–4] — and see Section 8 below).[83] Tarski also influenced model theory a great deal through his students and through his collaborations with other logicians.[84]

[79] In [Tarski, 1957b], Tarski and Vaught write:

> "A proof, along these lines, of Theorem 2.2 is known, but is by no means simple. It is essentially the same proof which was originally found by Tarski, in 1928, for the generalized Löwenheim-Skolem theorem." [Tarski, 1957b, p. 666, fn. 8]

[80] Vaught refers the reader to his speculations in [Vaught, 1965, p. 398], and [Vaught, 1974, p. 160]. The first published proof of (ii) appeared in [Mal'cev, 1936] (which also contains a generalization of Gödel's completeness and compactness theorems to uncountable languages).

[81] [Vaught, 1986, p. 870]. Two models are elementarily equivalent iff every sentence that is true in one is true in the other, and vice versa.

[82] This claim is made by [Chang and Keisler, 1973, p. 3], and by [Vaught, 1986, p. 876].

[83] For a more detailed survey of Tarski's postwar contributions to model theory, see [Vaught, 1986] and [Chang and Keisler, 1973, pp. 515–531]. Addison, Henkin and Tarski [1965b] contains a large bibliography of work on model theory up to that time.

[84] Consider the roster of logicians that Vaught provides:

5 DECIDABILITY AND UNDECIDABILITY

Tarski casts the decision problem sometimes in terms of proof, and sometimes in terms of truth. In [Tarski, 1953], for example, the notion of proof is central:

> "By a *decision procedure* for a given formalized theory T we understand a method which permits us to decide in each particular case whether a given sentence formulated in the symbolism of T can be proved by means of the devices available in T (or, more generally, can be recognized as valid in T). The *decision problem* for T is the problem of determining whether a decision procedure for T exists (and possibly of exhibiting such a procedure). A theory is called *decidable* or *undecidable* according as the solution of the decision problem is positive or negative." [Tarski, 1953, p. 3]

In [Tarski, 1948a], the notion of truth takes centre-stage:

> "The most important kind of decision problems is that in which K [a class of sentences] is defined to be the class of true sentences of a certain theory. When we say that there is a decision method for a certain theory, we mean that there is a decision method for the class of true sentences of the theory." [Tarski, 1948a, p. 1]

Tarski was in no doubt about the significance of the decision problem:

> "As is well known, the decision problem is one of the central problems of contemporary metamathematics". [Tarski, 1953, p. 3]

In the present section, we turn to Tarski's work on decidability (Sections 5.1 and 5.2) and undecidability (Section 5.3).

5.1 Decidability of the theory of the reals

Tarski famously proved the decidability of the first-order theory of the real numbers. The proof was first published in full detail in *A Decision Method for Elementary Algebra and Geometry* [Tarski, 1948a]:

> "In this monograph we present a method ... for deciding on the truth of sentences of the elementary algebra of real numbers." (p. 2)

"Tarski influenced model theory not only with his papers but also through his PhD students, his correspondence, and his conversations with people, many of whom came to Berkeley to see him. This was especially so in the two decades after the war. His PhD students active in model theory included Mostowski, Julia Robinson, Wanda Szmielew, [Robert Vaught], C. C. Chang, S. Feferman, Montague, H. J. Keisler, H. Gaifman, W. Hanf, and others. Some people (not PhD students) active in model theory, who were close to Tarski and received inspiration from him in their work, were Beth, Fraïssé, Henkin, Łoś, Lyndon, and D. Scott. One could add J. Ax, W. Craig, A. Ehrenfeucht, Y. Ershov, S. Kochen, M. Rabin, A. Robinson, and many, many others." [Vaught, 1986, p. 877]

By "the elementary algebra of real numbers" is to be understood

> "that part of the general theory of the real numbers in which one uses exclusively variables representing real numbers, constants denoting elementary operations on and relations between real numbers, like '+', '·', '−', '<', '>', and '=', and expressions of elementary logic such as 'and', 'or', 'not', 'for some x', and 'for all x'." (p. 2)

The primitive operations are '+' and '·', and so we shall refer to this theory as $\langle R, +, \cdot \rangle$.

Here decidability is understood in the semantic sense — the intuitive notion of truth plays a fundamental role in Tarski's presentation. (Tarski refers the reader to the formal definition in [Tarski, 1933a].) But Tarski points out that truth could be eliminated from the entire discussion by proceeding axiomatically, and replacing the notion of truth by the notion of provability (see [Tarski, 1948a, fn. 9, pp. 48–50], where Tarski provides a list of axioms). Under this new interpretation, Tarski's results lead easily to the conclusion that the axiomatic system of elementary algebra is decidable, consistent, and complete (see fn. 15, p. 53). Moreover, since the axioms are satisfied not only by the reals but also by the elements of any real-closed field, it follows that the theory of real-closed fields is also decidable, consistent, and complete (see fn. 15, p. 54).[85] Of these various related results, the decision method for real-closed fields is particularly celebrated.

Tarski established the decidability of the first-order theory of the reals by "the method of eliminating quantifiers" [Tarski, 1948a, p. 15]. Following [Tarski, 1948a, p. 50, fn. 11], we can trace the history of this method through [Löwenheim, 1915, Section 3], [Skolem, 1919, Section 4], [Langford, 1927; 1928], and [Presburger, 1930]. Tarski mentions that in his university lectures for the years 1926–8 the method was developed in a general and systematic way (see [Tarski, 1948a, p. 50, fn. 11]).

In outline, the method is as follows. Given a theory T, we identify certain *basic formulas*, and prove that every formula of T is T-equivalent to a Boolean combination of basic formulas. A *Boolean combination* of basic formulas is a formula obtained from basic formulas by repeated applications of $\neg$ and $\&$; and two formulas ϕ and ψ are *T-equivalent* iff $T \vDash \phi \leftrightarrow \psi$ (i.e. $\phi \leftrightarrow \psi$ is a semantic consequence of T). The proof — and in particular the step where we 'eliminate quantifiers' — yields a decision procedure for the theory T.

For illustration, consider the theory D of dense simple order without endpoints. We take D to be given by the set of its axioms,[86] and we take the true sen-

[85] For a full characterization of the theory of real-closed fields see, for example, [Chang and Keisler, 1973, p. 41].

[86] The theory of dense simple order without endpoints has the following axioms:

1. $\forall x \forall y \forall z (x \leq y \ \& \ y \leq z \rightarrow x \leq z)$ (transitivity)
2. $\forall x \forall y (x \leq y \ \& \ y \leq x \rightarrow x = y)$ (antisymmetry)
3. $\forall x (x \leq x)$ (reflexivity)
4. $\forall x \forall y (x \leq y \vee y \leq x)$ (comparability)

tences of D to be its axioms and their semantic consequences. We will apply the method of quantifier elimination to show that D is decidable (a result published by Langford in 1927).[87] We take as basic formulas the following: '$v_m = v_n$' and '$v_m \leq v_n$'. (Notice that the Boolean combinations of basic formulas are exactly the open formulas.) Let '$v_m < v_n$' abbreviate '$v_m < v_n$ & $v_m \neq v_n$'. We consider $n+1$ variables $v_0, v_1, \ldots, v_n$ for $n > 0$. We associate with these variables certain conjunctions θ_0 & θ_1 & $\ldots \theta_{n-1}$ called *arrangements*. Each conjunct θ_i is either $u_i < u_{i+1}$ or $u_i = u_{i+1}$, where $u_0, u_1, \ldots, u_n$ is a renumbering of $v_0, v_1, \ldots, v_n$.

The main result to be proved is the following:

Quantifier Elimination Theorem for D. Every formula is D-equivalent to an open formula.

We can easily show that to prove this theorem it is sufficient to prove:

(1) If $\psi(v_0, \ldots v_n)$ is an open formula, then $\exists v_m \psi$ is D-equivalent to an open formula.[88]

5. $\forall x \forall y (x \leq y \;\&\; x \neq y) \rightarrow \exists z (x \leq z \;\&\; z \neq x \;\&\; z \leq y \;\&\; z \neq y)$
6. $\exists x \exists y (x \neq y)$
7. $\forall x \exists y (x \leq y \;\&\; x \neq y)$
8. $\forall x \exists y (y \leq x \;\&\; x \neq y)$.

(Axioms 1,2,3 give the theory of *partial orders*; if we add Axiom 4, we obtain the theory of *simple order* (or *linear order*); if we further add Axioms 5 and 6, we obtain the theory of *dense simple order*; and with the final addition of Axioms 7 and 8, we arrive at the theory of dense simple order without endpoints.)

[87] In his papers [1927; 1928], Langford proved that the following theories are decidable by the method of quantifier elimination: dense linear orders without endpoints, dense linear orders with a first element but no last, dense linear orders with first and last elements, and the system of the natural numbers ordered by '<'.

[88] Here is a sketch of a proof of this result, that to prove the theorem it is sufficient to prove (1).

First we establish a general result for any theory T (this is Lemma 1.5.1 in [Chang and Keisler, 1973, p. 50]):

> To show that every formula is T-equivalent to a Boolean combination of basic formulas, it is sufficient to show:
>
> (i) every atomic formula is T-equivalent to a Boolean combination of basic formulas, and
>
> (ii) if ψ is a Boolean combination of basic formulas then $\exists v_m \psi$ is equivalent to a Boolean combination of basic formulas.

This general result is proved by induction on the complexity of formulas. If ϕ is an atomic formula, then ϕ is T-equivalent to a Boolean combination of basic formulas by (i). The cases for negation and conjunction are obvious. If ϕ is $\exists v_m \psi$, where ψ is a Boolean combination θ of basic formulas, then $\exists v_m \psi$ is T-equivalent to $\exists v_m \theta$, and so, by (ii), T-equivalent to a Boolean combination of basic formulas. (For more details, see Chang and Keisler, *ibid.*).

Now apply this general result to theory D. Here, (i) is immediate: the atomic formulas are $v_m = v_n$ and $v_m \leq v_n$, and these are the basic formulas (and so vacuously equivalent to a Boolean combination of basic formulas). So it remains only to prove (ii). Since the open (or quantifier-free) formulas are exactly the Boolean combinations of basic formulas, (ii) is equivalent to:

If ψ is an open formula, then $\exists v_m \psi$ is D-equivalent to an open formula.

As we are about to see, the proof of (1) involves the elimination of quantifiers, and yields a decision procedure for the theory D.[89] First, some preliminaries. Note that if $m > n$, then the quantifier is vacuous, and (1) is immediate. So we may assume $m \leq n$. It is convenient to further assume a suitable renaming of the variables so that $m = n$. We will also assume the following lemma:

Lemma. Every open formula $\psi(v_0, \ldots, v_n)$ is D-equivalent to $v_0 < v_0$, or $v_0 = v_0$, or the disjunction of finitely many arrangements of the variables $v_0, \ldots, v_n$.[90]

So for the purposes of proving (1), we can assume that ψ is

(i) $v_0 < v_0$, or

(ii) $v_0 = v_0$, or

(iii) the disjunction of finitely many arrangements of the variables $v_0, \ldots, v_n$.

(1) is immediate for the cases (i) and (ii), since in each case $\exists v_n \psi$ (either $\exists v_0(v_0 < v_0)$ or $\exists v_0(v_0 = v_0)$) is D-equivalent to the open formula ψ (either $v_0 < v_0$ or $v_0 = v_0$ respectively).

So we turn to case (iii), and let

$$\psi = \theta_0 \vee \ldots \vee \theta_p$$

where each θ_i is an arrangement of $v_0, \ldots, v_n$. It follows that

$$D \vDash \exists v_n \psi \leftrightarrow \exists v_n(\theta_0 \vee \ldots \vee \theta_p),$$

and so

$$D \vDash \exists v_n \psi \leftrightarrow \exists v_n \theta_0 \vee \ldots \vee \exists v_n \theta_p.$$

We now show that $\exists v_n \psi$ is D-equivalent to an open formula. First, suppose that $n = 1$. Then there are only three possibilities for each disjunct $\exists v_n \theta_i$: $\exists v_1(v_0 < v_1)$ or $\exists v_1(v_0 = v_1)$ or $\exists v_1(v_1 < v_0)$. Each of these is a consequence of D, and so the entire disjunction is a consequence of D. That is:

$$D \vDash \exists v_n \theta_0 \vee \ldots \vee \exists v_n \theta_p.$$

It follows that

$$D \vDash \exists v_n \psi.$$

So $\exists v_n \psi$ is D-equivalent to $v_0 = v_0$, an open formula.

Second, suppose that $n > 1$. We now describe a procedure for finding an open formula D-equivalent to $\exists v_n \psi$. (We will later see that this procedure is the core of

But this is just (1) — and so it remains only to prove (1). So to prove the quantifier elimination theorem it is sufficient to prove (1) — which is what we wanted to show.

[89] For the proof that follows, I follow Chang and Keisler [Chang and Keisler, 1973, pp. 52–4]. Full details can be found there.

[90] This is Lemma 1.5.2, [Chang and Keisler, 1973, p. 51]. For the proof, see pp. 51–2.

the decision procedure for D.) Consider each arrangement θ_i of $v_0, \ldots, v_n$. There is an associated arrangement $\theta_i{}^*$ of $v_0, \ldots, v_{n-1}$ formed from θ_i by leaving out v_n.[91] It is straightforward to check that

$$D \vDash \exists v_n \theta_i \leftrightarrow \theta_i^*.^{92}$$

So it follows that

$$D \vDash \exists v_n \psi \leftrightarrow \theta_0^* \vee \ldots \vee \theta_p^*.$$

So $\exists v_n \psi$ is D-equivalent to an open formula. This completes the proof of (1), and so we have proved the quantifier elimination theorem for D.

We are also in a position to describe a decision procedure for D. Given an arbitrary sentence ϕ, we wish to determine whether ϕ is a true sentence of D. We proceed as follows. We put ϕ into prenex normal form $\mathrm{Q}_0 v_0 \ldots \mathrm{Q}_n v_n \psi$, where each Q_i is a quantifier expression, and ψ is an open formula. We can assume without loss of generality that Q_n is an existential quantifier, since otherwise we may start with $\neg\phi$. We now eliminate the existential quantifier Q_n. Since ψ is an open formula, by the above lemma it is D-equivalent to $v_0 < v_0$ or $v_0 = v_0$ or a disjunction $\theta_0 \vee \ldots \vee \theta_p$ of finitely many arrangements of $v_0, \ldots, v_n$. In the first case, we replace $\mathrm{Q}_n v_n \psi$ by $v_0 < v_0$; in the second case we replace $\mathrm{Q}_n v_n \psi$ by $v_0 = v_0$. In the third case, we replace $\mathrm{Q}_n v_n \psi$ by $v_0 = v_0$ if $n = 1$, and if $n > 1$, we replace $\mathrm{Q}_n v_n \psi$ by a D-equivalent open formula $\theta_0^* \vee \ldots \vee \theta_p^*$ by the procedure just described in the previous paragraph. We repeat this procedure for $\mathrm{Q}_{n-1}, \mathrm{Q}_{n-2}, \ldots, \mathrm{Q}_1$, until a sentence of the form $\mathrm{Q}_0 v_0 \chi(v_0)$ remains. And then it is obvious whether or not $\mathrm{Q}_0 v_0 \chi(v_0)$ is a true sentence of D. (It is easy to see that $\mathrm{Q}_0 v_0 \chi(v_0)$ must take one of four forms: $\exists v_0(v_0 = v_0), \forall v_0(v_0 = v_0), \exists v_0(v_0 < v_0)$ or $\forall v_0(v_0 < v_0)$, the first two of which are clearly true sentences of D, while the second two clearly are not.) From this we can decide whether or not ϕ is a true sentence of D.

For a simple example, consider the sentence

(i) $\exists v_0 \forall v_1 \exists v_2 (v_0 < v_1 \;\&\; v_1 < v_2)$.

[91] For example, suppose $n = 2$ and the arrangement θ_i of v_0, v_1 and v_2 is:

$$v_0 < v_1 \;\&\; v_1 = v_2.$$

Then the associated arrangement θ_i^* of v_0 and v_1 is:

$$v_0 < v_1.$$

[92] For an illustration, consider θ_i and θ_i^* from the previous footnote. Here the biconditional $\exists v_n \theta_i \leftrightarrow \theta_i^*$ is

$$\exists v_2 (v_0 < v_1 \;\&\; v_1 = v_2) \leftrightarrow v_0 < v_1.$$

which is logically equivalent to

$$v_0 < v_1 \;\&\; \exists v_2 (v_1 = v_2) \leftrightarrow v_0 < v_1.$$

This is clearly a semantic consequence of D as required (since $\exists v_2(v_1 = v_2)$ is an obvious semantic consequence of D).

Notice that this sentence is already in prenex normal form, and that the formula '$v_0 < v_1$ & $v_1 < v_2$' is already an arrangement of the variables v_0, v_1, v_2. Following our procedure, we replace the formula '$\exists v_2(v_0 < v_1$ & $v_1 < v_2)$' by '$v_0 < v_1$', and obtain the sentence:

(ii) $\exists v_0 \forall v_1(v_0 < v_1)$.

Since the innermost quantifier expression of (ii) is a universal quantifier, we work instead with its negation, which is logically equivalent to

(iii) $\forall v_0 \exists v_1 \neg(v_0 < v_1)$.

Since here $n = 1$, our procedure tells us to replace '$\exists v_1 \neg(v_0 < v_1)$' by '$v_0 = v_0$' to obtain the sentence

(iv) $\forall v_0(v_0 = v_0)$.[93]

We readily see that (iv) and hence (iii) are true sentences of D. And since (iii) is logically equivalent to the negation of (ii), we conclude that (ii) and hence (i) are not true sentences of D.

This completes our illustration of a decision procedure generated by the proof of a quantifier elimination theorem. Tarski's decision procedure for the first-order theory $\langle R, +, \cdot\rangle$ follows the same general pattern. The central result of [Tarski, 1948a] is a quantifier elimination theorem for $\langle R, +, \cdot\rangle$ (Theorem 31, p. 39).[94] The decision method is a straightforward consequence of the theorem, and follows soon after (Theorem 37, p. 42).

The proofs and results of [Tarski, 1948a] have a long history. Tarski reports that he had obtained partial results tending in the same direction — such as the decidability of elementary algebra with addition as the only operation, and of the geometry of the straight line — in his university lectures of 1926–8 (see [Tarski, 1948a, fn. 4], where Tarski refers the reader to to Presburger [1930, p. 95, fn. 4], and Tarski [1921, p. 324, fn. 3].) Tarski remarks that the decision procedure for the first-order theory of the reals was found in 1930 (see the Preface of [Tarski, 1948a] and also p. 2, especially fn. 4). Tarski's paper [Tarski, 1931a] on the definability of the reals (which we have discussed in Section 4 above) implicitly mentions a

[93]Notice how this step connects to the case of $n = 1$ in the proof above. The formula '$\neg(v_0 < v_1)$' is D-equivalent to the following disjunction of arrangements of v_0, v_1:

$$v_0 = v_1 \vee v_1 < v_0.$$

Now $\exists v_1(v_0 = v_1 \vee v_1 < v_0)$ is logically equivalent to

$$\exists v_1(v_0 = v_1) \vee \exists v_1(v_1 < v_0).$$

It is clear that each disjunct is a logical consequence of D. And so we may replace '$\exists v_1 \neg(v_0 < v_1)$' by '$v_0 = v_0$'.

[94]Tarski remarks that, from a purely mathematical point of view, the results leading up to Theorem 31, and Theorem 31 itself, are closely related to Sturm's theorem. Indeed, Tarski writes that "Theorem 31 constitutes an extension of Sturm's theorem... to arbitrary systems of equations and inequalities with arbitrarily many unknowns." (p. 52, fn. 12).

quantifier elimination result for $\langle R, 1, \leq, + \rangle$, where the basic formulas are '$x = 1$', '$x \leq y$', '$x + y = z$'. Recall Theorem 1 from [Tarski, 1931a] (mentioned in Section 4.1 above):

Theorem. A set S of sequences of real numbers is a member of Df if and only if S is a finite sum of finite products of elementary linear sets.

Theorem 1 is a theorem of mathematics, but as we noted in Section 4.1 there is a correlation between (mathematical) sets of sequences and (metamathematical) sentential functions, and between their respective Boolean operations — and so there is a metamathematical analogue of Theorem 1. Tarski writes:

> We can easily formulate (and prove) a metamathematical theorem which is an exact analogue of Th. 1. This metamathematical result leads us to a conclusion that, in the system of arithmetic described in Sec 1 [viz. $\langle R, 1, \leq, + \rangle$], every sentence of order 1 can be proved or disproved. Moreover, by analysing the proof of this result, we see that there is a mechanical method which enables us to decide in each particular case whether a given sentence (of order 1) is provable or disprovable. [Tarski, 1931a, p. 134]

The metamathematical result that Tarski mentions here is a quantifier elimination theorem for $\langle R, 1, \leq, + \rangle$ expressed in terms of definability. Tarski also mentions that the decidability of $\langle R, 1, \leq, + \rangle$ (as well as its completeness) follows from the proof of the theorem.

Immediately after the passage just quoted, Tarski points out that the theory of the real numbers can be based on different systems of primitive concepts — for example, it can be based on the concepts of sum and product. Then a modified form of Theorem 1 holds for $\langle R, +, \cdot \rangle$. Here we can see an implicit mention of the quantifier elimination theorem for $\langle R, +, \cdot \rangle$, since the modified version of Theorem 1 will have its metamathematical analogue too.

Tarski first *explicitly* mentions the quantifier elimination theorem for $\langle R, +, \cdot \rangle$ in *The completeness of elementary algebra and geometry* [Tarski, 1967], a monograph that was scheduled for publication in 1940.[95] Tarski also provides an outline of the proof. In this work the emphasis is on the *completeness* of $\langle R, +, \cdot \rangle$ (see [Tarski, 1967, Theorem 2.1, Sec. 2]). Referring to [Tarski, 1967] and [Tarski, 1948a], Tarski writes:

[95] The monograph was to appear in 1940 in *Actualites Scientifiques et Industrielles*, Hermann & Cie, Paris. In the Foreword to [Tarski, 1967], Tarski reports that the process of publication reached the stage of page proofs, but then, during the war, the whole setting was destroyed. All that remained were two sets of page proofs in Tarski's possession. Several years later, J. C. C. McKinsey re-wrote the monograph, and this version was published as [Tarski, 1948a]. In 1967, the original page proofs were reprinted, largely due to the efforts of René de Possel, Director of the Institut Blaise Pascal in Paris. Sczcerba remarks (in [Szczerba, 1986, p. 908]) that Tarski first saw the reprint of the original proofs in 1967 on his 65th birthday.

> "A comparison of the titles of the two monographs reveals that the center of scientific interest has been shifted to the decision problem from that of completeness". (Foreword to [Tarski, 1967])

Indeed, from [Tarski, 1931a] to [Tarski, 1967] and on to [Tarski, 1948a] we see a shift in emphasis from definability to completeness to decidability. In [Tarski, 1948a], mention of definability and completeness is relegated to footnotes. In fn. 13 of [Tarski, 1948a] Tarski remarks that the quantifier elimination theorem "gives us a simple characterization of those sets of real numbers, and relations between real numbers, which are arithmetically definable", and refers the reader back to [Tarski, 1931a]. Completeness is mentioned in fn. 15. The emphasis on decidability is at least in part explained by the fact that the RAND corporation supported the publication of [Tarski, 1948a]. Tarski writes:

> "Within a few months the monograph was published. As was to be expected, it reflected the specific interests which the RAND Corporation found in the results. The decision method for elementary algebra and geometry — which is one of the main results of the work — was presented in a systematic and detailed way, thus bringing to the fore the possibility of constructing an actual decision machine. Other, more theoretical aspects of the problems discussed were treated less thoroughly, and only in notes." (Preface to [Tarski, 1948a])

5.2 The decidability of elementary geometry

As Tarski mentions here, and as its title suggests, the monograph [Tarski, 1948a] contains a proof not only of the decidability of the first-order theory of the reals but also of the decidability of elementary (that is, first-order) geometry. Tarski outlines a decision method for the specific case of 2-dimensional Euclidean geometry, but points out that the method can be adapted to Euclidean geometry of any number of dimensions, as well as to various systems of non-Euclidean and projective geometry.[96] Tarski observes in fn. 18 (pp. 55–57) that, just as in the case of elementary algebra, one may proceed axiomatically, in terms of provability rather than truth. Tarski presents a list of axioms for 2-dimensional Euclidean geometry, which is easily modified to form a basis for elementary geometry of any number of dimensions. One of the fruits of the decision method in this axiomatic setting is a "constructive consistency proof for the whole of elementary geometry" (fn. 18, p. 57).[97]

The decision method for 2-dimensional Euclidean geometry is readily obtained from the decision method for elementary algebra, since each sentence of elementary geometry can be suitably correlated with a sentence of elementary algebra.

[96] In fn. 19, Tarski mentions that ordinary projective geometry may be treated as a branch of the theory of modular lattices, referring the reader to Birkhoff [Birkhoff, 1948]. Tarski goes on to say that the decision method also applies to this branch of lattice theory.

[97] Tarski observes that this improves a result to be found in [Hilbert and Bernays, 1939, vol 2, pp. 38ff].

Following Tarski (pp. 43–45), we will first describe the system of 2-dimensional Euclidean geometry and then show how to set up the correlation.

The language of 2-dimensional Euclidean geometry contains infinitely many variables ranging over points of the Euclidean plane, and three predicate constants: the identity sign '=', the 3-place predicate '$B(x, y, z)$' to be read as 'y is between x and z', and the 4-place predicate '$D(x, y, z, w)$' to be read as 'the distance from x to y is equal to the distance from z to w'. The atomic formulas are '$x = y$', '$B(x, y, z)$', and '$D(x, y, z, w)$'; complex formulas are built from the atomic formulas via the usual sentential connectives and the quantifiers. It is noteworthy that in Tarski's formalization only points are treated as individuals, and there are no set-theoretical devices or second-order variables. This is in contrast to Hilbert's influential *Grundlagen der Geometrie*, where certain geometrical figures are treated as individuals, and the others are treated as second-order point sets.[98]

We can translate a sentence ϕ of elementary geometry into a sentence ϕ^* of elementary algebra by fixing a coordinate system. We make the following replacements:

(i) If x is a (geometric) variable in ϕ, we replace it by two (algebraic) variables x_1 and x_2. (Intuitively, x represents a point in the Euclidean plane, and x_1 and x_2 represent its real coordinates.) Distinct geometric variables are replaced by distinct algebraic variables.

(ii) The quantifier phrases $\exists x$ and $\forall x$ are replaced by $\exists x_1 \exists x_2$ and $\forall x_1 \forall x_2$ respectively.[99]

(iii) '$x = y$' is replaced by '$x_1 = y_1$ & $x_2 = y_2$'.

(iv) '$B(x, y, z)$' is replaced by

$$'(y_2 - x_2) \cdot (z_1 - y_1) = (z_2 - y_2) \cdot (y_1 - x_1) \;\&\; 0 \leq (x_1 - y_1) \cdot (y_1 - z_1) \;\&\; 0 \leq (x_2 - y_2) \cdot (y_2 - z_2)'.$$

[98] Tarski draws the contrast in [Tarski, 1959]

> "Thus, in our formalization of elementary geometry, only points are treated as individuals and are represented by (first-order) variables. Since elementary geometry has no set-theoretical basis, its formalization does not provide for variables of higher orders and no symbols are available to represent or denote geometrical figures (point sets), classes of geometrical figures, etc. It should be clear that, nevertheless, we are able to express in our symbolism all the results which can be found in textbooks of elementary geometry and which are formulated there in terms referring to various special classes of geometrical figures, such as the straight lines, the circles, the segments, the triangles, the quadrangles, and, more generally, the polygons with a fixed number of vertices, as well as to certain relations between geometrical figures in these classes, such as congruence and similarity. This is primarily a consequence of the fact that, in each of the classes just mentioned, every geometrical figure is determined by a fixed finite number of points." [Tarski, 1959, p. 17, and see also fn. 3]

[99] Alternatively, we can work just with '$\exists$', as Tarski does, and take '$\forall$' to abbreviate '$\neg\exists\neg$'.

(v) '$D(x, y, z, w)$' is replaced by

$$'(x_1 - y_1)^2 + (x_2 - y_2)^2 = (z_1 - w_1)^2 + (z_2 - w_2)^2'.$$

For example, the true sentence

$$\forall x \forall y \forall z \exists w D(x, y, z, w)$$

of elementary geometry is correlated with the true sentence

$$\forall x_1 \forall x_2 \forall y_1 \forall y_2 \forall z_1 \forall z_2 \exists w_1 \exists w_2 [(x_1 - y_1)^2 + (x_2 - y_2)^2 = (z_1 - w_1)^2 + (z_2 - w_2)^2]$$

of elementary algebra. Tarski makes the general observation:

> It is now obvious to anyone familiar with the elements of analytic geometry that whenever ϕ is true then ϕ^* is true, and conversely. [Tarski, 1948a, p. 45]

It follows that 2-dimensional elementary geometry is decidable:

> And since we can always decide in a mechanical way about the truth of ϕ^*, we can also do this for ϕ. (*ibid.*)

Tarski proved a number of other decidability results that we have not yet mentioned, all by the method of quantifier elimination. They include decidability results for the theory of dense linear orderings (proved in Tarski's lectures of 1926–8); for the theory of wellorderings (work started with Mostowski in the 1930s, and culminating in [Tarski, 1978]); and for the theory of Boolean algebras (announced in [Tarski, 1949f]).[100] Although the method of quantifier elimination did not originate with Tarski, it was in his hands an extremely powerful systematic tool, producing a number of important results about a wide range of theories. None were more significant than the decision method for real-closed fields, described by Doner and Hodges as "an astonishingly fruitful mathematical result" [Doner and Hodges, 1988, p. 23]; for a detailed account of the impact of this result of Tarski's, see van den Dries [van den Dries, 1988, esp. pp. 10–16].

5.3 Undecidability

In his paper "A general method in proofs of undecidability" [Tarski, 1953, pp. 3–35],[101] Tarski distinguishes two different methods for solving the decision problem.

[100] For more details and further references, see Doner and Hodges [Doner and Hodges, 1988, pp. 21–23]. Several of Tarski's students — Presburger, Szmielew, and Doner — proved decidability results under Tarski's supervision (see [Doner and Hodges, 1988, pp. 23–4]). In particular, Presburger proved the decidability of the arithmetic of the natural numbers with addition in [Presburger, 1930] — his master's thesis, supervised by Tarski.

[101] This paper is the first of three that compose [Tarski, 1953] (written in collaboration with Mostowski and A. Robinson). Tarski reports that the observations contained in the paper were made in 1938–9, presented to a meeting of the Association for Symbolic Logic in 1948, and summarized in [Tarski, 1949b].

1. *The direct method.* The 'direct' method is based on ideas developed in [Gödel, 1931], in particular the notions of general recursive functions and sets, and the following characterisation of a decidable theory:

> "A theory T is called *decidable* if the set of all its valid sentences is recursive, and otherwise *undecidable*." [Tarski, 1953, p. 14]

This is the method used by Church to show the undecidability of Peano's arithmetic, and by Rosser to show that every consistent extension of Peano's arithmetic is also undecidable.

2. *The indirect method.* This method

> "consists in reducing the decision problem for a theory T_1 to the decision problem for some other theory T_2 for which the problem has previously been solved." [Tarski, 1953, p. 4]

In the original form of the indirect method, there are two ways to proceed:

(i) Where T_2 is undecidable, we show that T_1 can be obtained from T_2 by deleting finitely many of T_2's axioms.

(ii) Where T_2 is essentially undecidable, we show that T_2 is interpretable in T_1.

Church applied procedure (i) to prove that first-order predicate logic is undecidable, taking T_2 to be a fragment of Peano arithmetic (PA). Tarski observes that procedure (ii) often proceeds with PA as T_2 — in this way, for example, "various axiomatic systems of set theory have turned out to be undecidable" (p. 4).

However, Tarski also observes that the indirect method is limited:

> "The indirect method in its original form was rather restricted in applications. Only in exceptional cases can a theory for which the decision problem is discussed be obtained from another theory, which is known to be undecidable, simply by omitting finitely many sentences from the axiom system of the latter. On the other hand, one could hardly expect to find an interpretation of Peano's arithmetic in various simple formalized theories, with meager mathematical contents, for which the decision problem was open. With regard to theories of this kind both the direct and indirect methods seem to fail." (p. 4)

Tarski's novel contribution was to extend and modify the indirect method, significantly enlarging the scope of its application. Here the key notion is that of an *essentially undecidable theory*: a theory is *essentially undecidable* if it and every consistent extension of it is undecidable. The observation at the heart of Tarski's method is this:

(I) "[I]n order to establish the undecidability of a theory T_1, it suffices to show that some essentially undecidable theory T_2 can be interpreted, not necessarily in T_1, but (what is much easier) in some consistent extension of T_1 — provided only that T_2 is based upon a finite axiom system." (p. 5)

Since T_2 must be finitely axiomatizable, PA can no longer serve as T_2. However,

> "examples of essentially undecidable theories which are based upon finite axiom systems and are readily interpretable in other theories have been found (by the direct method) among fragments of Peano's arithmetic. Using this fact and applying the extended indirect method, many formalized theories — like the elementary theories of groups, rings, fields, and lattices — have recently been shown to be undecidable." (p. 5)

A particularly useful example is the theory Q, a finitely axiomatizable and essentially undecidable fragment of PA found by Tarski and Mostowski using the direct method in 1939 (see [Tarski, 1949c]), and considerably simplified by Robinson (in [Robinson, 1950]).[102] Tarski writes:

> "Theory Q turned out to be very suitable for our method; its mathematical content is meager, and it can easily be interpreted or at least weakly interpreted in many different theories. Hence Theory Q has become a powerful instrument in the study of the decision problem; with its help it has proved to be possible to obtain a negative solution of this problem for a large variety of theories for which the problem had previously been open." [Tarski, 1953, p. 32]

In [Tarski, 1953], Tarski lays out the theoretical foundations for his indirect method. In sections I.2 and I.3, Tarski defines a number of basic concepts and proves six theorems about undecidable and essentially undecidable theories, with no mention yet of the notion of interpretability. We first review some of Tarski's definitions:

> "A theory T_1 is called a *subtheory* of a theory T_2 if every sentence which is valid in T_1 is also valid in T_2; under the same conditions T_2 is referred to as an *extension* of T_1." (p. 11)
>
> "An extension T_2 of T_1 is called *inesssential* if every constant of T_2 which does not occur in T_1 is an individual constant and if every valid sentence of T_2 is derivable in T_2 from a set of valid sentences of T_1." (p. 11)
>
> "Two theories T_1 and T_2 are said to be *compatible* if they have a common consistent extension." (p. 12)
>
> A theory is *axiomatizable* "if there is a recursive set A of valid sentences of T such that every valid sentence of T is derivable from the set A; if the set A is assumed to be finite, the theory T is called *finitely axiomatizable*." (p. 14)

[102] The construction of Q is carried out in "Undecidability and Essential Undecidability in Arithmetic", by Mostowski, Robinson and Tarski, the second paper in [Tarski, 1953]. For historical information about the construction, see [Tarski, 1953, pp. 39–40, fn. 1].

One useful theorem Tarski proves in section I.3 is the following:

Theorem. Let T_1 and T_2 be two compatible theories such that every constant of T_2 is also a constant of T_1. If T_2 is essentially undecidable and finitely axiomatizable, then T_1 is undecidable, and so is every subtheory of T_1 which has the same constants as T_1. (This is Theorem 6, p. 18.)

Later in the paper (p. 32), Tarski gives an example of an application of Theorem 6. Consider theory N, the arithmetic of the natural numbers, with non-logical constants + and ·, and possibly others, such as 0 and 1 and <. By definition, a sentence of N is valid if it is true under the standard intepretation. Let us observe the following:

(i) N is undecidable, and since it is also complete, it is essentially undecidable and not axiomatizable.[103]

(ii) PA is an axiomatic subtheory of N with the same constants, and though axiomatizable, PA is not finitely axiomatizable.

(iii) In turn, Q is a subtheory of PA which is finitely axiomatizable, and essentially undecidable.

Given (iii), PA has a finitely axiomatizable, essentially undecidable subtheory. And so it follows from our observations that N has a finitely axiomatizable, essentially undecidable subtheory. Now if we apply Theorem 6

> "we arrive at once at the conclusion that every theory which is compatible with Q and has the same constants as Q is undecidable. Hence, in particular, every subtheory of N in which the set of constants includes + and · is undecidable ... this generalizes the results known from the literature which concern various special subtheories of N." (p. 32)

The theorems Tarski proves in section 3 — including Theorem 6 — make no mention of interpretability. In section 4, Tarski widens the scope of these theorems by introducing the notion of interpretability. It is here in section 4 that Tarski explicitly provides the grounds for the method encapsulated in (I) above. The notion of intepretability rests on the prior notion of *definition*, or more specifically the notion of a *possible definition of a given constant in a theory T*. Tarski introduces the notion by way of examples, one of which is the 2-place predicate symbol <:

> "A possible definition of < in T is any sentence of the form
>
> (i) $\forall x \forall y(x < y \leftrightarrow \phi)$

[103] Tarski states the following theorem in [Tarski, 1953, p. 14]:

> For a complete theory T the following three conditions are equivalent: (i) T is undecidable, (ii) T is essentially undecidable, and (iii) T is not axiomatizable.

The proof is straightforward: for a complete theory, (i) implies (iii), and the other parts of the theorem are immediate from the definitions.

> where ϕ is a formula of T (x and y being any two different variables such that no variable different from both of them occurs free in ϕ)." (p. 20)

Notice that while (i) is not a sentence of T, it is a sentence in every extension of T which contains $<$ as a constant. It is straightforward to extend the notion of a possible definition to n-place predicate and operation letters.

Now consider two theories T_1 and T_2, and assume for the moment that T_1 and T_2 have no non-logical constants in common. Tarski characterizes the notion of interpretablity of one theory in another as follows:

> "... T_2 is *interpretable* in T_1 if we can extend T_1, by including in the set of valid sentences some possible definitions of the nonlogical constants of T_2, in such a way that the resulting extension of T_1 turns out to be an extension of T_2 as well." (p. 21)

The notion of interpretability is easily extended to the general case in which T_1 and T_2 may have some non-logical constants in common.[104] Tarski also introduces the notion of *weak interpretability*:

> "A theory T_2 is said to be *weakly interpretable* in T_1 if T_2 is interpretable in some consistent extension of T_1 which has the same constants as T_1." (p. 21)

Tarski goes on in section I.4 to prove several theorems that turn on the notions of interpretability and weak interpretability. One of these theorems is the following:

Theorem. Let T_1 and T_2 be two theories such that T_2 is weakly interpretable in T_1 or in some inessential extension of T_1. If T_2 is essentially undecidable and finitely axiomatizable, then:

(i) T_1 is undecidable and every subtheory of T_1 which has the same constants as T_1 is undecidable;

(ii) there exists a finite extension of T_1 which has the same constants as T_1 and is essentially undecidable. (Theorem 8, pp. 23–4)

Theorem 8 is a generalization of Theorem 6, for if T_1 and T_2 are compatible and every constant of T_2 is also a constant of T_1, then T_2 is weakly interpretable in T_1. Tarski remarks that, from the point of view of providing theoretical foundations for his general indirect method,

[104] In the general case

> "we first replace the non-logical constants in T_2 by new constants not occurring in T_1 (different symbols by different symbols), without changing the structure of T_2 in any other respect; if the resulting theory T_2' proves to be interpretable in T_1, we say that T_2 is also interpretable in T_1." (p. 21)

> "Theorem 8 is especially important ...; for, given a finitely axiomatizable and essentially undecidable theory, this theorem enables us to establish the undecidability of various other theories which may be very distant in their mathematical content from the original theory." (p. 30)

Theorem 8 captures in a precise way the observation in (I) above.

For an example of the application of theorems from section 4 (including Theorem 8), we need one more notion, that of *relative interpretability*. Given a theory T and a 1-place predicate P, the theory $T^{(P)}$ is obtained by relativizing the quantifiers in T to P; that is, each subformula $\forall x\psi$ (or $\exists x\psi$) is replaced by $\forall x(Px \rightarrow \psi)$ (or $\exists x(Px\&\psi)$). Now, a theory T_2 is *relatively interpretable* (*relatively weakly interpretable*) in a theory T_1 if $T_2^{(P)}$ is interpretable (weakly interpretable) in T_1. Among the theorems concerning relative interpretability is the following:

Theorem. Let T be any theory and P a unary predicate which is not a constant of T. Then $T^{(P)}$ is essentially undecidable if and only if T is essentially undecidable. (Theorem 10, p. 27)

Now consider again the theories N and Q, and note that, by Theorem 10, $Q^{(P)}$ is essentially undecidable. Tarski writes:

> "N is known to be relatively interpretable in the arithmetic of integers J, a theory constructed analogously to N, but in which the variables are assumed to range over the set of all integers. Hence, Q is also relatively interpretable in J. By applying Theorems 8–10 we conclude that every subtheory of J in which $+, \cdot$, and possibly some other symbols occur as nonlogical constants is undecidable. As interesting examples of such subtheories we may mention the elementary theories of rings, commutative rings, and ordered rings." (p. 33)

A parallel result was established by Julia Robinson for the arithmetic A of rational numbers (see [Robinson, 1949]). Robinson proved that N, and so Q, is relatively interpretable in A — and that consequently all the subtheories of A (with $+$ and $\cdot$ as non-logical constants) are undecidable. These subtheories include the elementary theories of fields and ordered fields.

As Tarski mentions (pp. 33–34), many other results were established using the indirect method, by Tarski and by others.[105] Tarski established the undecidability of the elementary theories of groups, lattices, modular lattices, complemented modular lattices, and abstract projective geometries (see [Tarski, 1949d; Tarski, 1949e]). Tarski also established undecidability results for equational the-

[105] For a survey of more recent results inspired by Tarski's method, see [McNulty, 1986, esp. pp. 893–6].

ories of relation algebras,[106] and for certain branches of elementary geometry.[107] W. Szmielew and Tarski proved that all known axiomatic systems of set theory, with the 1-place predicate 'S' ('is a set') and '$\in$' ('is a member of') as nonlogical constants, are undecidable; they proved this general result by showing that Q is interpretable in a small axiomatic, essentially undecidable fragment of set theory.[108] R. M. Robinson showed that the elementary theories of various special rings are undecidable (see [Robinson, 1951]). Grzegorczyk proved the undecidability of distributive lattices, Brouwerian algebras, and related algebraic and geometric systems (see [Grzegorczyk, 1951]).

In his proofs of decidability results, Tarski took a known method — the elimination of quantifiers — to a new level. But in his proofs of undecidability results, Tarski established a new method, an indirect method that turned on the semantic notions of *definition* and *interpretability*. Indeed, Tarski's method is often referred to as the "method of interpretation". So at the heart of Tarski's original contributions to the decision problem lies his groundbreaking work on fundamental semantic notions.

6 MORE ON LOGIC AND ALGEBRA

Tarski's work on algebra and logic flows in two directions, from algebra to logic, and back again. As we saw in 2.4 above, Tarski increasingly 'algebraized' logic, investigating logical systems as interpretations of algebras. In the other direction, Tarski brought metalogical questions to bear on algebras; we saw, for example, that Tarski established the decidability of Boolean algebras (5.1 above) and the undecidability of equational theories of relation algebras (5.3 above). In this section, we say more about Tarski's work on the connections between algebra and logic.

6.1 Boolean algebra and the calculus of classes

With the first direction in mind, recall that in [Tarski, 1935+1936] Tarski treated the calculus of sentences and the calculus of deductive systems as two realizations of Boolean algebra (see 2.4 above). In his 1935 paper "On the Foundations of Boolean Algebra" [Tarski, 1935], Tarski showed that another logical theory, the calculus of classes, can be treated in the same way.[109] Tarski opens his paper as

[106] These results can be found in [Tarski, 1987a, section 8.5, pp. 251–258]. The history of these results is reviewed in section 8.7, pp. 268–271. Relation algebras are characterized on pp. 235–236. As we will see in Section 6 of this chapter, Tarski initiated the study of relation algebras in his paper [Tarski, 1941b]. An equational theory is a restricted elementary theory, restricted to those universal sentences whose quantifier-free subformulas are equations between terms.

[107] See [Tarski, 1959] and [Tarski, 1979].

[108] These and related results are stated without proof in [Tarski, 1952b].

[109] We presented the calculus of classes in 4.2 above (it was the object language for Tarski's definition of truth in [Tarski, 1933a]). Variables range over classes, and there are four constants: negation, disjunction, the universal quantifier, and inclusion.

follows:

> "Boolean algebra, also called the *algebra of logic*, is a formal system with a series of important interpretations in various fundamental departments of logic and mathematics. The most important and best known interpretation is the *calculus of classes*." [Tarski, 1935, p. 320]

Tarski goes on to investigate several systems of Boolean algebra, and shows that one of them — the 'atomistic system of Boolean algebra' — is equivalent with the calculus of classes. So, along with [Tarski, 1935+1936], [Tarski, 1935] is a contribution to the algebraization of logic.

Recall Postulates I–VII (presented in 2.4 above, pp. 523–524). These form what Tarski calls the *ordinary system of Boolean algebra*. To these axioms, Tarski adds three more, each of an infinite character. These axioms contain two new primitive operations: $\sum_{y\in X} y$ (the sum of all elements of the set X) and $\prod_{y\in X} y$ (the product of all elements of the set X). The additional axioms are as follows (where, recall, B is the universe of discourse):

Postulate VIII If $X \subseteq B$, then
(a) $\sum_{y\in X} y \in B$;
(b) $x < \sum_{y\in X} y$ for every $x \in X$;
(c) if $z \in B$ and $x < z$ for every $x \in X$, then $\sum_{y\in X} y < z$.

Postulate IX If $X \subseteq B$, then
(a) $\prod_{y\in X} y \in B$;
(b) $\prod_{y\in X} y < x$ for every $x \in X$;
(c) if $z \in B$ and $z < x$ for every $x \in X$, then $z < \prod_{y\in X} y$.

Postulate X If $x \in B$ and $X \subseteq B$, then
(a) $x.\sum_{y\in X} y = \sum_{y\in X}(x.y)$;
(b) $x + \prod_{y\in X} y = \prod_{y\in X}(x + y)$.

Postulates I–X form the *extended* (or *complete*) *system of Boolean algebra*.[110]

In the second half of [Tarski, 1935], Tarski constructs the *atomistic system of Boolean algebra*. The set of atoms, At, is defined as follows:

Definition. $x \in At$ if and only if (1) $x \in B$ and $x \neq 0$, and (2) for every element $y \in B$, the formulas $y < x$ and $y \neq 0$ imply $y = x$.

Tarski now adds one further axiom to Postulates I–X:

Postulate D If $x \in B$ and $x \neq 0$, then there is an element $y \in At$ such that $y < x$.

[110] In the first half of [Tarski, 1935], Tarski shows that this axiomatization can be simplified — first to a system of four axioms, and then to a system of two (see pp. 323–333). Tarski reports (in a footnote added later to [Tarski, 1935, on p. 333]) that these simplifications were influenced by work of Leśniewski, specifically the deductive theory that Leśniewski called *mereology* (see Leśniewski [1916] and [1927–31]).

Without Postulate D, the following questions are all open: Do atoms exist? Are atoms included in every element (other than 0)? Is every element the sum of the atoms included in it? With Postulate D, all these questions are answered in the affirmative. The system of postulates I–X+D is the *atomistic system of Boolean algebra*. It is this system of algebra that Tarski shows to be equivalent with the calculus of classes (see Theorem 6, p. 340).[111]

6.2 The calculus of relations and relation algebras

In a paper published a few years later — "On the Calculus of Relations" [Tarski, 1941b] — Tarski shows that we can take the algebraic view of another logical theory, the *calculus of (binary) relations*. Tarski traces the beginnings of this theory back to de Morgan, but "[t]he title of creator of the theory of relations was reserved for C. S. Peirce" [Tarski, 1941b, p. 73]. Peirce's work "was continued and extended in a very thorough and systematic way by E. Schröder" (*ibid.*) in *Algebra und Logik der Relative* ([Schröder, 1895]).[112] Schröder's thorough account of the calculus of relations contains a large number of unsolved problems, and indicates the direction for further investigations — consequently, Tarski expresses his amazement that there was no specific development of this rich logical theory in the decades that followed. Tarski sought "to awaken interest in a certain neglected logical theory" [Tarski, 1941b, p. 89].

Tarski contrasts two methods for constructing the calculus of relations. The first constructs the calculus as part of a larger logical theory (the functional calculus developed by Hilbert and Ackermann). The second method is specific to the calculus of relations, and as a result it is simpler and more elegant. This method is an algebraic construction which originates with Tarski, and we turn to it now.

The vocabulary of Tarski's construction comprises relation variables, the usual sentential connectives, the familiar constants and operations from Boolean algebra ($1, 0, ^-, +$, and $\cdot$), and five further symbols peculiar to the calculus of relations: $1'$ (the identity element, or Peircean unit), $0'$ (the diversity element, or Peircean

[111] Tarski remarks that Postulate D "brings with it a whole series of far-reaching consequences" [Tarski, 1935, p. 335]. Among these are the general distributive laws for addition and multiplication:

(E) Let R be any class of subsets X of the set B ($\sum_{X\in R} X \subseteq B$), and let S be the class of all those sets Y which are included in the sum of all sets of the class R and have at least one element in common with every set of the class R ($Y \subseteq \sum_{X\in R} X$, and if $X \in R$, then $X.Y \neq 0$). Then:

(a) $\prod_{X\in R} \sum_{z\in X} z = \sum_{Y\in S} \prod_{z\in Y} z$, and

(b) $\sum_{X\in R} \prod_{z\in X} z = \prod_{Y\in S} \sum_{z\in Y} z$.

The sentences (a) and (b) are generalizations of Postulates V and X respectively. Tarski and Lindenbaum showed that (E) cannot be derived from Postulates I–X alone; indeed, given Postulates I–X, (E) is equivalent to Postulate D.

Tarski's work on general distributive laws in Boolean algebras, started here in [Tarski, 1935], was developed further in [Tarski, 1957a].

[112] This work of Schröder's was the third volume of *Vorlesungen über die Algebra der Logik* [Schröder, 1890–1905].

zero), $\breve{}$ (the unary operation of conversion), ; (the binary operation of relative product), and $\dagger$ (the binary operation of relative addition).[113] To fix ideas, think for the moment of relations set-theoretically, as sets of ordered pairs. Then the identity element is the set of identity pairs, the converse of a relation R is given by:

$$\breve{R} = \{\langle x, y\rangle | \langle y, x\rangle \in R\},$$

and the relative product of R and S is given by

$$R; S = \{\langle x, z\rangle | \exists y \text{ such that } \langle x, y\rangle \in R \ \& \ \langle y, z\rangle \in S\}.$$

Peircean zero and relative addition are dispensable, since each can be defined in terms of the other symbols (as we shall see shortly).

The axioms of Tarski's algebra fall into three groups: the usual axioms for the sentential connectives, the axioms for Boolean algebra (with class variables replaced by relation variables), and the axioms governing the further 'Peircean' constants and operations. Tarski lists eight axioms in this third group:

(i) $\breve{\breve{R}} = R.$

(ii) $\overbrace{\mathrm{R;S}} = \breve{S}; \breve{R}.$

(iii) $R; (S; T) = (R; S); T$

(iv) $R; 1' = R.$

(v) $R; 1 = 1 \vee 1; \bar{R} = 1.$

(vi) $(R; S).\breve{T} = 0 \rightarrow (S; T).\breve{R} = 0.$

(vii) $0' = \bar{1}'.$

(viii) $R \dagger S = \overline{\bar{R}; \bar{S}}.$

Axioms (i)–(iv) involve exclusively $1', \breve{}$, and ;. Though fully abstract, they are natural and obvious when we read them in the more concrete setting of set theory. Axioms (v) and (vi) establish connections between Boolean and 'Peircean' concepts. Axioms (vii) and (viii) can be understood as definitions of $0'$ and $\dagger$. [114]

[113] Perhaps more familiar now than $1', 0', ;$, and $\dagger$ are the symbols $\mathring{1}$, $\mathring{0}$, $\odot$, and $\oplus$, respectively. See for example, [Tarski, 1987a, pp. 235–6], where a characterization of a *relation algebra* may also be found.

[114] Tarski points out that these axioms can be represented geometrically. Suppose that the relation variables denote relations between real numbers, and consider a rectangular coordinate system in the plane. Then every relation R may be represented as the set of all points (x, y) such that x has the relation R to y. The relations $1, 0, 1', 0'$ are respectively: the whole plane, the empty set of points, the identity relation represented by the straight line with equation $x = y$, and the diversity relation by the set of all points not on this straight line. The Boolean operations are represented in the usual set-theoretic way. To obtain $\breve{R}$, take the set of points — or point

Here we have the first presentation of the calculus of relations as an axiomatic algebraic theory. Moreover, Tarski's paper [Tarski, 1941b] initiated the study of *relation algebras*. Tarski's work on the arithmetic of relation algebras at Berkeley during the 1940s culminated with the paper "Distributive and modular laws in the arithmetic of relation algebras" [Tarski, 1951a], which Tarski wrote with his student Louise Chin. In [Tarski, 1951b] and [Tarski, 1952a], Parts 1 and 2 of "Boolean algebras with operators", Tarski and Jonsson investigated algebraic systems enriched by new operations, and relation algebras figure prominently in this work.[115]

Tarski's paper [Tarski, 1941b] is another clear example of Tarski's algebraic treatment of logical theories. But it also shows Tarski working in the reverse direction, raising metalogical questions for his algebraic theory. For example, Tarski raises a question as to the equivalence of the two methods of construction. Declaring this a difficult open question, Tarski continues:

> "I can only say that I am practically sure that I can prove with the help of the second method all of the hundreds of theorems to be found in Schröder's *Algebra und Logik der Relative*." (p. 88)

For a second example, Tarski turns to the so-called "representation problem":

> "Is every model of the axiom system of the calculus of relations isomorphic with a class of binary relations which contains the relations

set — corresponding to R and rotate it (in three dimensions) through an angle of 180^o about the line $x = y$. It is rather more complicated to explain relative product geometrically. We need to supplement the coordinate system with a z-axis perpendicular to the xy-plane. Then for the representation of $R; S$ (in the xy-plane) proceed as follows: rotate the point set corresponding to R through a right angle about the x-axis, draw through every point of the resulting set a straight line parallel to the y-axis, and take the union of all these straight lines — this gives us the 'cylindrical' point set R^*. Similarly, rotate the set corresponding to S through a right angle about the y-axis, draw the lines parallel to the x-axis, and obtain the 'cylindrical' point set S^*. Finally, take the intersection of R^* and S^* and project it orthogonally on the xy-plane. This projection is the geometrical representation of $R; S$. The representation of relative sum is readily obtained from that of the relative product, given axiom (viii).

This geometric representation provides an intuitive reading of the axioms. For example, axiom (i) corresponds to the obvious geometric fact that if we rotate any point set through 180^o about a given straight line, and then again by 180^o about the same straight line, we obtain the original point set. Or consider axiom (v): we can see that this axiom is true under the geometric representation. Observe first that $R; 1$ holds between x and y if and only if there is a z such that xRz (the relation is independent of y). So to obtain the geometric representation of $R; 1$ consider the set corresponding to R, draw through every point of this set a straight line parallel to the y-axis, and take the union of all these straight lines. Now axiom (v) says that either the set representing $R; 1$ is the whole plane or the set representing $1; \bar{R}$ is the whole plane. If $R; 1$ is not the whole plane then there must be a line parallel to the y-axis which does not contain any point in the set corresponding to R. So this line is composed only of points belonging to the set corresponding to $\bar{R}$. If through every point of this line we draw a straight line parallel to the x-axis and take the union of all these parallel lines, we obtain the whole plane. That is, the relation $1; \bar{R}$ is represented by the whole plane.

[115] For more on the work by Tarski and others on relation algebras, see [Monk, 1986, esp. pp. 901–2] and [Jonsson, 1986, esp. p. 884].

> $1, 0, 1', 0'$ and is closed under all the operations considered in this calculus?" (p. 88)

M. H. Stone had proved the analogous result for Boolean algebra in the affirmative (see [Stone, 1936]). In 1941 Tarski's question was an open one — but Lyndon proved that the answer is in the negative (see [Lyndon, 1950], with a correction in [Lyndon, 1956]). Tarski maintained his interest in the metalogic of relation algebras throughout the rest of his career — Tarski's undecidability results for a variety of relation algebras, mentioned above and in 5.2, are contained in his final published monograph [Tarski, 1987a] (with Givant).[116]

6.3 Predicate logic and cylindric algebras

Thus far we have seen that Tarski provides algebraic treatments of a number of logical theories: the calculus of sentences, the calculus of deductive systems, the calculus of classes, and the calculus of (binary) relations. We turn now to Tarski's algebraic treatment of the predicate calculus. To this end, Tarski developed the theory of *cylindric algebras*:

> "This theory ... was originally designed to provide an apparatus for an algebraic study of first-order predicate logic." [Tarski, 1971, Foreword, p. 1]

Tarski collaborated with Henkin and Monk to produce *Cylindric Algebras Part 1* [Tarski, 1971] and *Part 2* [Tarski, 1985].

In the Foreword to [Tarski, 1971], Tarski, Henkin and Monk introduce cylindric algebras by way of *cylindric set algebra*. Just as Boolean algebra is an abstraction from Boolean set algebra (see Section 2.4 above, p. 522), so cylindric algebra is an abstraction from cylindric set algebra. So let us begin with cylindric set algebras. In the broadest terms, a *cylindric set algebra of dimension* α is a structure

$$A = \langle S, \cup, \cap, \sim, 0, {}^{\alpha}U, C_\kappa, D_{\kappa\lambda} \rangle_{\kappa,\lambda<\alpha}$$

where S is the universe of the algebra (the set of all its elements), $\cup$ and $\cap$ are binary operations, $\sim$ and $C_\kappa(\kappa < \alpha)$ are unary operations, and $0, {}^{\alpha}U, D_{\kappa\lambda}(\kappa, \lambda < \alpha)$ are distinguished elements. We now describe these ingredients in more detail.

It is helpful to describe cylindric set algebras using terminology drawn from analytic geometry (see [Tarski, 1971, pp. 1–2]). We start with the set ${}^{\alpha}U$. The elements of ${}^{\alpha}U$ are sequences $x = \langle x_0, x_1, \ldots, x_k, \ldots \rangle$ of length α (α an ordinal), where each x_k is a member of an arbitrary set U. The set ${}^{\alpha}U$ is called the *α-dimensional Cartesian space with base U*, and each element x is called a *point*,

[116] Tarski discusses other metalogical problems in [Tarski, 1941b], specific to the first method of constructing the calculus of relations. He announces two results: (i) the calculus of relations is undecidable, and (ii) there are sentences of the elementary theory of relations with only relation variables free that cannot be transformed into an equivalent sentence of the calculus of relations. Tarksi leaves unsolved a new decision problem regarding the transformations mentioned in (ii): is there a procedure for deciding in a particular case whether such a transformation is possible?

with *coordinates* $x_0, x_1, \ldots, x_k, \ldots$. The elements of the universe S are subsets of ${}^{\alpha}U$. (Just as the elements of a Boolean set algebra may be subsets of an arbitrary set V, so the elements of a cylindric algebra are subsets of a certain Cartesian power of an arbitrary set U.) The operations $\cup$, $\cap$ and $\sim$ are the usual Boolean operations; indeed, the structure

$$A = \langle S, \cup, \cap, \sim, 0, {}^{\alpha}U \rangle$$

is a Boolean set algebra, with S closed under the Boolean operations.

The distinguished element $D_{\kappa\lambda}$ and the unary operation C_κ are specific to cylindric set algebras. $D_{\kappa\lambda}$ is a certain subset of ${}^{\alpha}U$ — its members are the points of ${}^{\alpha}U$ whose κth coordinate is identical to its λth coordinate. The point set $D_{\kappa\lambda}$ is called the κ, λ-*diagonal set*. The aptness of the name is easily seen if we consider the case $\alpha = 2$ (the case where the Cartesian space with the base U is 2-dimensional) — here $D_{\kappa\lambda}$ is the main diagonal line of the coordinate system. In general, for $\kappa \neq \lambda, D_{\kappa\lambda}$ is the hyperplane defined by $x_\kappa = x_\lambda$.

The unary operation C_κ is called the κth *cylindrification*. Intuitively, given a subset X of the space ${}^{\alpha}U$, $C_\kappa(X)$ is the cylinder obtained if each point in X is 'stretched out' parallel to the κth coordinate axis. More precisely,

> a point y is in $C_\kappa(X)$ iff there is a point in X which differs from y only in its κth coordinate.

(It is easy to check that if C_κ is applied to the diagonal set $D_{\kappa\lambda}$, it yields ${}^{\alpha}U$.)[117] The universe S is closed under c_κ.

We now move to the general notion of a *cylindric algebra* by abstracting from cylindric set algebras. We consider the algebraic identities that hold in *all* cylindric set algebras, and select some of them as axioms for the general, abstract theory of cylindric algebras. Changing the terminology, we take a cylindric algebra of dimension α to be a structure

$$\mathfrak{A} = \langle A, +, \cdot, -, 0, 1, c_\kappa, d_{\kappa\lambda} \rangle_{\kappa,\lambda<\alpha}$$

where A is an arbitrary set, $+, \cdot, -, 0, 1$ are the familiar Boolean operations and distinguished elements, c_κ (for $\kappa < \alpha$) is a unary operation (still called the *κth cylindrification*), and $d_{\kappa\lambda}$ (for $\kappa, \lambda < \alpha$) is a distinguished element (called the *$\kappa\lambda$-diagonal element*). The axioms of a cylindric algebra fall into two groups. The first group of axioms characterize the Boolean algebra $\langle A, +, ., -, 0, 1 \rangle$. (For example, we could use Postulates I–VII from 2.4 above.) The second group is composed of equational axioms that are special to cylindric algebras. These axioms are as follows, where x, y are arbitrary elements of A, and κ, λ, μ are arbitrary ordinals less than α:

1. $c_\kappa(0) = 0$

[117] In general, $C_\kappa(D_{\kappa\lambda}) = C_\lambda(D_{\kappa\lambda}) = {}^{\alpha}U$. For example, consider the case $\alpha = 2$. If the 01 diagonal is stretched out parallel to either the 0 axis or the 1 axis, the result is the entire 2-dimensional Cartesian space with base U.

2. $x \cdot c_\kappa(x) = x$
3. $c_\kappa(x \cdot c_\kappa(y)) = c_\kappa(x) \cdot c_\kappa(y)$
4. $c_\kappa(c_\lambda(x)) = c_\lambda(c_\kappa(x))$
5. $c_\kappa(d_{\kappa\lambda}) = 1$
6. $c_\mu(d_{\kappa\mu} \cdot d_{\mu\lambda}) = d_{\kappa\lambda}$, provided that $\mu \neq \kappa, \lambda$
7. $c_\kappa(d_{\kappa\lambda} \cdot x) \cdot c_\kappa(d_{\kappa\lambda} \cdot -x) = 0$, provided $\kappa \neq \lambda$

These axioms are natural enough if we keep in mind cylindric set algebra.[118] But in order to grasp the motivation for them, and for cylindric algebras generally, we must remember their original purpose: to serve "as an instrument for the algebraization of predicate logic" [Tarski, 1971, p. 4]. Cylindric algebras can be viewed not only as an abstraction from cylindric set algebra, but also from first-order logic:

> "The notion of a cylindric algebra can be considered as a common, algebraic abstraction from its two sources." [Tarski, 1985, p. v]

So consider first-order predicate logic with identity, together with Tarski's definitions of *model* and (semantic) *consequence*. Let Φ be the denumerable set of all formulas of (first-order) predicate logic. Let Σ be a set of sentences — or a *theory* — from the language of predicate logic. We now define the notion of equivalence relative to a theory Σ as follows:

Definition. ϕ and ψ are equivalent under set Σ, written $\phi \equiv_\Sigma \psi$, iff $\phi \leftrightarrow \psi$ is a consequence of Σ.

It is easy to see that $\equiv_\Sigma$ is an equivalence relation on Φ. So for each formula ϕ, there is an equivalence class $[\phi]$ of all formulas ψ such that $\phi \equiv_\Sigma \psi$, and the union of these mutually exclusive equivalence classes is Φ (that is, $\equiv_\Sigma$ partitions Φ). Let Φ_Σ be the set of these equivalence classes.

Now we form a *cylindric algebra for a theory* Σ:

$$\mathfrak{A}_\Sigma = \langle \Phi_\Sigma, +, \cdot, -, 0, 1, c_\kappa, d_{\kappa\lambda} \rangle_{\kappa,\lambda<\omega}$$

where

$$\begin{aligned} [\phi] + [\psi] &= [\phi \vee \psi] \\ [\phi] \cdot [\psi] &= [\phi \& \psi] \\ -[\phi] &= [\neg\phi] \\ c_\kappa[\phi] &= [\exists v_\kappa \phi] \\ d_{\kappa\lambda} &= [v_\kappa = v_\lambda] \end{aligned}$$

[118] For example, consider Axiom 2 in the concrete setting of cylindrical set algebra. It is obvious that if we consider the set of points x and its cylindrification $c_\kappa(x)$, then the points they share in common are just the points of x. And we have already seen the intuitive backing for Axiom 5.

and the distinguished elements 0 and 1 are $[v_\kappa \neq v_\kappa]$ and $[v_\kappa = v_\kappa]$ respectively. It is straightforward to check that this algebra satisfies all the axioms of ω-dimensional cylindric algebras, in particular Axioms 1–7.[119]

The significance of cylindrical algebras for the algebraic study of predicate logic "is rather obvious" [Tarski, 1971, p. 8]. To paraphrase what Henkin, Monk and Tarski go on to say, theories are basic entities in metalogical discussions just as algebraic structures are basic entities in algebraic research. By correlating the algebra $\mathfrak{A}_\Sigma$ with any theory Σ, we have established a correspondence (which turns out to be one-one) between first-order theories and cylindric algebras of formulas. The basic metalogical problems for a fixed theory Σ are problems of the type: "Is a given sentence ϕ a consequence of the theory Σ?". *Each such problem clearly reduces to an algebraic problem concerning the associated algebra of formulas*: "Does the equation $[\phi]$=1 hold in the algebra $\mathfrak{A}_\Sigma$?"

For Tarski, then, cylindric algebras (like Boolean and relational algebras) were a tool for algebraizing logic. But Tarski also brought logic to bear on the algebras — like Boolean and relational algebras, cylindrical algebras are themselves a fit subject for metalogical investigation. Chapter 4 of [Tarski, 1985] applies logical notions to cylindrical algebras: the authors consider the model theory of cylindrical algebras in the first part of the chapter, and various decision problems in the second. Decidability and undecidability results are obtained for various cylindrical algebras.

In the Foreword to [Tarski, 1971], Tarski, Henkin and Monk trace the path that led to their collaboration:

> "The theory of cylindric algebras was founded by Tarski, in collaboration with his former students Louise H. Chin (Lim) and Frederick B. Thompson, during the period 1948-52. Soon thereafter Henkin became interested in the subject and began to work with Tarski on its further development. In 1961 they published a fairly extensive outline of their research, and the plan was first formulated to prepare a detailed monograph on the subject. Subsequently Monk's substantial contributions to the theory made a joining of efforts desirable, and thus the present team of authors was finally formed." [Tarski, 1971, p. 23]

Part 2 of *Cylindric Algebras* [Tarski, 1985] was published in 1985; its first draft was two-thirds complete when Tarski died. The monograph *Cylindric set algebras* [Tarski, 1981], authored by Tarski, with Andreka, Henkin, Monk and Nemeti, was another major contribution to the theory. The theory of cylindric algebras has been an active field of research: for a survey of major results and further references, see [Monk, 1986, esp. 903–5].

[119] For example, Axiom 1 holds since the existential generalization of any member of $[v_\kappa \neq v_\kappa]$ is a member of $[v_\kappa \neq v_\kappa]$; Axiom 2 holds since $[\phi] = [\phi \& \exists x_\kappa \phi]$; Axiom 4 turns on the equivalence of $\exists v_\kappa \exists v_\lambda \phi$ and $\exists v_\lambda \exists v_\kappa \phi$; and Axiom 5 holds since $c_\kappa(d_{\kappa\lambda}) = [\exists v_\kappa (v_\kappa = v_\lambda)] = [v_\kappa = v_\kappa] = 1$.

7 MORE ON GEOMETRY

7.1 The metamathematics of elementary geometry

We have already seen one of Tarski's most celebrated metamathematical result in geometry: the decidability of elementary (or first-order) Euclidean geometry, discovered in 1930 but not published in full until [Tarski, 1948a] (see section 5.2 above). Tarski returned to the metamathematics of geometry in "What is elementary geometry?" [Tarski, 1959]. Recall that in [Tarski, 1948a, fn. 18] Tarski briefly considers an axiomatic theory of Euclidean geometry — it is this theory of elementary geometry that Tarski investigates further in [Tarski, 1959].

In [Tarski, 1959], Tarski is fully explicit about the reference of the term 'elementary geometry'. Loosely it refers to the geometry which is based on Euclid's *Elements* and which forms the subject matter of secondary school geometry. But for metamathematical investigations, we need a more precise description; for one thing, we must specify the language in which the sentences of elementary geometry are formulated. As we saw in Section 5.2 above, the language of elementary geometry is the *first-order predicate calculus* — the first-order variables range over individuals (points), and there are no second-order or set-theoretical devices. The language contains two primitive predicates: a three-place predicate $Bxyz$ ('y is between x and z'), and the 4-place predicate $Dxyzw$ ('the distance from x to y is equal to the distance from z to w').

Tarski goes on to present a list of 12 axioms and one axiom schema. (See [Tarski, 1959, pp. 19–20]. This list supersedes the one in [Tarski, 1948a, fn. 18], which was found to contain superfluous axioms. Such refinements aside, Tarski's axiom system essentially dates back to his university lectures in the years 1926–27, as Tarski reports in [Tarski, 1967, p. 341, fn. 34]. These axioms form the basis for elementary geometry, or G for short. They are the universal closures of the following:

A1. $Bxyx \rightarrow x = y$. (Identity Axiom for Betweenness)

A2. $Bxyu \ \& \ Byzu \rightarrow Bxyz$. (Transitivity Axiom for Betweenness)

A3. $Bxyz \ \& \ Bxyu \ \& \ x \neq y \rightarrow Bxzu \vee Bxuz$. (Connectivity Axiom for Betweenness)

A4. $Dxyyx$. (Reflexivity Axiom for Equidistance)

A5. $Dxyzz \rightarrow x = y$. (Identity Axiom for Equidistance)

A6. $Dxyzu \ \& \ Dxyvw \rightarrow Dzuvw$. (Transitivity Axiom for Equidistance)

A7. $Bxzw \ \& \ Byuz \rightarrow \exists v(Bwvy \ \& \ Bxuv)$. (Pasch's Axiom)[120]

[120] In [Tarski, 1979], Tarski and Sczcerba provide the following figure for Pasch's axiom:

A8. $Bxut \;\&\; Byuz \;\&\; x \neq u \rightarrow \exists v \exists w(Bxyv \;\&\; Bxzw \;\&\; Bvtw)$. (Euclid's Axiom)[121]

A9. $Dxyx'y' \;\&\; Dyzy'z' \;\&\; Dxux'u' \;\&\; Dyuy'u' \;\&\; Bxyz \;\&\; Bx'y'z' \;\&\; x \neq y \rightarrow Dzuz'u')$. (Five-Segment Axiom)

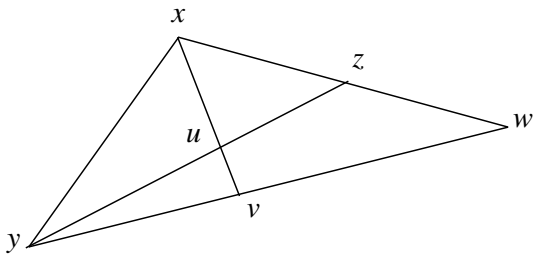
A10. $\exists z(Bxyz \;\&\; Dyzuv)$. (Axiom of Segment Construction)

A11. $\exists x \exists y \exists z(\neg Bxyz \;\&\; \neg Byzx \;\&\; \neg Bzxy)$. (Lower Dimension Axiom)

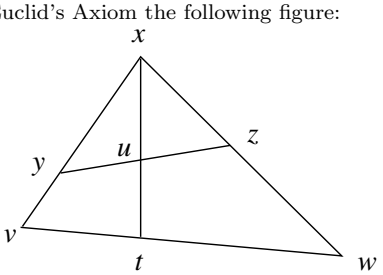
A12. $Dxuxv \;\&\; Dyuyv \;\&\; Dzuzv \;\&\; u \neq v \rightarrow Bxyz \vee Byzx \vee Bzxy$. (Upper Dimension Axiom)

A13. (Elementary Continuity Axioms)

$$\exists z \forall x \forall y(\phi \& \psi \rightarrow Bzxy) \rightarrow \exists u \forall x \forall y(\phi \& \psi \rightarrow Bxuy)$$



[121] We can associate with Euclid's Axiom the following figure:



A8 says in essence: there is always a straight line through a point t inside an angle (here, angle yxz) that touches both sides of the angle (see [Tarski, 1983b, p. 13]).

Notice that the line vw meets the lines yv and zw in such a way that the angles yvt and twz add up to less than two right angles. Compare A8 with Euclid's original formulation of his Fifth Postulate:

> "That, if a straight line [vw] falling on to straight lines [yv and zw] make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet [at x] on that side on which are the angles less than two right angles."

A8 is provably equivalent to the Parallel Axiom (Axiom IV in [Hilbert, 1922]):

> If A is a line and a is a point that does not belong to A, then there is exactly one line which is parallel to A to which a belongs.

For a proof of the Parallel Axiom (Hilbert's IV) from A8, see [Tarski, 1983b, p. 123]

where ϕ and ψ are formulas, z and y do not occur free in ϕ, and z and x do not occur free in ψ.

Axiom 8 is crucial to the formation of Euclidean geometry. If we omit it, we have the theory of absolute geometry; and if we admit instead its negation, we have Bolyai–Lobachevsky geometry. The system comprising A1–A12 is worthy of study in its own right, since much of Euclidean geometry proceeds without A13.[122]

A13 is a first-order analogue of the second-order axiom of continuity. Tarski notes that standard axiomatizations of G include a continuity axiom that contains second-order variables $X, Y, \ldots$ ranging over sets of points. A typical formulation is:

$$\forall X \forall Y [\exists z \forall x \forall y (x \in X \ \& \ y \in Y \rightarrow Bzxy) \rightarrow \exists u \forall x \forall y (x \in X \ \& \ y \in Y \rightarrow Bxuy)].$$

The first-order schema contained within A13 is obtained by replacing '$x \in X$' by an arbitrary first-order formula in which x occurs free, and '$y \in Y$' by an arbitrary first-order formula in which y occurs free. A13 can be viewed as a restriction of the second-order axiom to the definable sets.

Tarski goes on to address four fundamental metamathematical problems. The first is the *representation problem*. In general, the representation problem for a theory is the problem of characterizing all of the theory's models. In the case of G, the answer is given by the following theorem:

Representation Theorem. M is a model of G if and only if M is isomorphic with the Cartesian space over some real closed field F.[123]

The second and third problems are the *completeness problem* and the *decidability problem*. Tarski provides positive answers to both: G is complete and consistent [Tarski, 1959, p. 22, Theorem 2], and decidable [Tarski, 1959, p. 23, Theorem 3]. The fourth problem is the *problem of finite axiomatizability*. Here the question is: Can the axiom system be replaced by an equivalent finite system? The answer is negative: G is not finitely axiomatizable [Tarski, 1959, p. 24, Theorem 4].[124]

[122] For more in this vein, see [Szczerba, 1986, p. 909].

[123] This is Theorem 1 in [Tarski, 1959, p. 21]. A model of G is a structure $\langle A, \mathrm{B}, \mathrm{D}\rangle$, where A is an arbitrary non-empty set, B is a 3-place relation and D is a 4-place relation, and where all the axioms of G hold if the variables are taken to range over elements of A, and the constants B and D are taken to refer to B and D respectively. Let F be a real closed field $\langle F, +, \cdot, =\rangle$. (For a definition of *real closed field*, see for example [Chang and Keisler, 1973, p. 41].) Consider the set $A_F = F \times F$ of all ordered pairs $x = \langle x_1, x_2\rangle$ with x_1 and x_2 in F. We define the relations B_F and D_F as follows:

$$\begin{array}{lll} \mathrm{B}_F(x,y,z) & \text{iff} & (x_1 - y_1)\cdot(y_2 - z_2) = (x_2 - y_2)\cdot(y_1 - z_1) \ \& \ 0 \leq (x_1 - y_1)\cdot(y_1 - z_1) \ \& \\ & & \quad 0 \leq (x_2 - y_2)\cdot(y_2 - z_2). \\ \mathrm{D}_F(x,y,z,w) & \text{iff} & (x_1 - y_1)^2 + (x_2 - y_2)^2 = (z_1 - w_1)^2 + (z_2 - w_2)^2. \end{array}$$

(Compare clauses (iv) and (v) in Section 5.2) The structure $C_2 = \langle A_F, \mathrm{B}_F, \mathrm{D}_F\rangle$ is called the *Cartesian space over* F. We now have all the ingredients of Theorem 1. Tarski provides an outline of a proof in [Tarski, 1959, pp. 21–22].

[124] Tarski provides an outline of the proof on pp. 23–24.

This negative result might be considered a drawback of Tarski's treatment of elementary geometry. Further, there are notions of textbook geometry that cannot be expressed in G — for example, the notions of the circumference and area of a circle, and the notion of a polygon with arbitrarily many vertices. In [Tarski, 1959], Tarski considers two other ways of interpreting the term 'elementary geometry' by way of alternative axiomatizations. The first is of greater expressive power than G, and can accommodate the geometric notions just mentioned; however, Tarski notes that this theory is undecidable, and that the other three metamathematical problems remain open. The second alternative is finitely axiomatizable; however, it is weaker than G and consequently captures fewer validities, and its decision problem is open. Tarski concludes:

> "The problem of deciding which of the various formal conceptions of elementary geometry is closer to the historical tradition and the colloquial usage of this notion seems to be rather hopeless and deprived of broader interest. The author feels that, among these various conceptions, the one embodied in [G] distinguishes itself by the simplicity and clarity of underlying intuitions and by the harmony and power of its metamathematical implications." (p. 28)

A recent, detailed study of Tarski's theory G can be found in [Tarski, 1983b]. This, and its bibliography give a good sense of Tarski's extensive influence on research in foundational geometry and metamathematics.

Tarski's interest in the metamathematics of geometry was not restricted to Euclidean geometry. As we noted in Section 5.2, Tarski proved that certain systems of non-Euclidean and projective geometry are decidable; and as we noted in Section 5.3, Tarski showed that abstract projective geometries are undecidable (see [Tarski, 1949e]). Tarski also investigated the metalogic of *affine geometry* in [Tarski, 1965a] and [Tarski, 1979], in collaboration with L.W. Szczerba. As Szczerba puts it, affine geometry is roughly the theory of the betweenness relation on the open and convex subsets of the Euclidean plane.[125] A precise axiomatic characterization of general affine geometry (GA) is presented in [Tarski, 1979]. The system of axioms for GA consists of A1, A2, A3, A7, A11, and A13, together with three further axioms specific to affine geometry.[126] Tarski and Szczerba solve the representation problem for GA, and go on to show that GA is incomplete (in fact GA has $2^{\aleph_0}$ complete extensions), undecidable (in fact hereditarily undecid-

[125] See [Szczerba, 1986, p. 910].

[126] The three further axioms are as follows:

Extension Axiom The universal closure of $\exists x(Bxyz \;\&\; x \neq y)$.

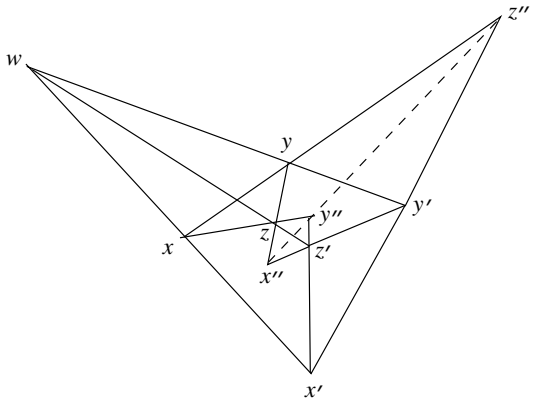
Desargues' Axiom The universal closure of $Bwxx' \;\&\; Bwyy' \;\&\; Bwzz' \;\&\; Bxyz'' \;\&\; Bx'y'z'' \;\&\; Byzx'' \;\&\; By'z'x'' \;\&\; Bxzy'' \;\&\; Bx'z'y'' \;\&\; \neg Bwxy \;\&\; \neg Bxyw \;\&\; \neg Bywx \;\&\; \neg Bwyz \;\&\; \neg Byzw \;\&\; \neg Bzwy \;\&\; \neg Bwzx \;\&\; \neg Bzxw \;\&\; \neg Bxwz \rightarrow Bx''y''z''$.

able),[127] and not finitely axiomatizable. In [Tarski, 1979] Tarski and Szczerba also investigate the metamathematical properties of various extensions of GA. One of these — Euclidean affine geometry (EA) — is obtained by adding Euclid's Axiom A8 to the axioms for GA. EA is the main subject of [Tarski, 1965a], where it is shown, like GA, to be incomplete, (hereditarily) undecidable, and not finitely axiomatizable.

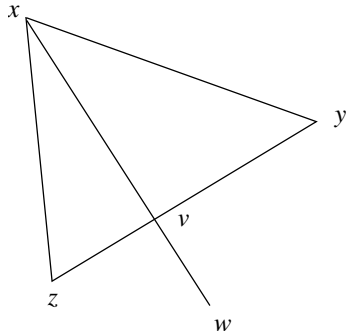
7.2 Primitive concepts of geometry

Another aspect of Tarski's metamathematical treatment of geometry is his work on primitive notions in geometry. We have seen that Tarski's formalization of Euclidean geometry requires just two primitive concepts — *betweenness* and *equidistance*. For Tarski, this economy is desirable and metamathematically significant. (In this respect, Tarski's formalization is superior to Hilbert's.) The question arises: How economical can we be? Can we find a single geometric notion in terms of which all the concepts of Euclidean geometry can be defined?

More narrowly, we can ask whether there is a relation between points that



Upper Dimension Axiom The universal closure of $\exists v[\{Byvz \ \& \ (Bxvw \vee Bvwx \vee Bwxv)\} \vee \{Bzvx \ \& \ (Byvw \vee Bvwy \vee Bwyv)\} \vee \{Bxvy \ \& \ (Bzvw \vee Bvwz \vee Bwzv)\}]$.



[127] A theory T is *hereditarily undecidable* if not only T but also every subtheory of T with the same constants is undecidable.

can serve this purpose. This is the leading question of Tarski's papers [Tarski, 1956a] (with E.W.Beth) and [Tarski, 1956b]. This search is constrained by a result established by Tarski and Lindenbaum: No *binary* relation between points can serve this purpose.[128] Still, Pieri showed that a certain three-place relation I comes close, where $I(x, y, z)$ holds iff the distance from x to y is the same as the distance from x to z (see [Pieri, 1908]). (In the 2-dimensional case, $I(x, y, z)$ holds iff x, y and z are the vertices of an isosceles triangle, with the line from y to z as the base; in the 1-dimensional case, x is the midpoint of the line from y to z.) The relation I *can* serve as the only primitive for n-dimensional Euclidean geometry, *provided that* $n \geq 2$. The equidistance relation denoted by D also comes close, since for $n \geq 2$ the betweenness relation B is definable in terms of D.

Can we do better than I and D? In [Tarski, 1956a], Tarski considers the 'equilaterality' relation $E : E(x, y, z)$ holds if the distances between x, y and z are equal. (In the 2-dimensional case, $E(x, y, z)$ holds if x, y and z are the vertices of an equilateral triangle.) Can E serve as the only primitive notion of Euclidean geometry? Tarski shows that the answer is affirmative for n-dimensional Euclidean geometry as long as $n \geq 3$; the answer is negative for $n = 1$ or $n = 2$. So E joins the ranks of other ternary relations, like I, which can serve as the sole primitive concept of n-dimensional Euclidean geometry for certain n, but not all n. In [Tarski, 1956b], Tarski considers in full generality the question of which ternary relations can serve as the sole primitive of Euclidean geometry for all n, and establishes a property that any such relation must have (and which E and I in particular fail to have, as Tarski shows).

Tarski's search for a single primitive relation in [Tarski, 1956a] and [Tarski, 1956b] was limited to relations between points. Much earlier, in the paper 'Foundations of the geometry of solids' [Tarski, 1929a], Tarski developed a geometry of solids, in which the primitive individuals are not points but spheres — or better, *balls*.[129] Tarski was responding to a challenge set by Leśniewski: to establish the foundations of a geometry of solids. In laying these foundations, Tarski works within the framework of Leśniewski's deductive theory of mereology (see Leśniewski, [1916] and [1927–31]). Just one 2-place relation — *the relation of part to whole* — is needed for Leśniewski's theory. Tarski showed in [Tarski, 1929a] that we can define all the concepts of 3-dimensional Euclidean geometry in terms of just the notions of *ball* and *the part-whole relation*.

The key to Tarski's demonstration is this: we may define the notion of *point* and

[128] See Lindenbaum and Tarski [1936c]. Indeed, Lindenbaum and Tarski announce that in elementary Euclidean geometry no 2-place relation can be defined "apart from the universal relation, the empty relation, identity and diversity." [Tarski, 1936c, p. 389]

[129] In the original paper [Tarski, 1929a], Tarski used the word for 'sphere'. But in a footnote added later (fn. 1, p. 26), Tarski suggests that, since he is developing a geometry of solids and not a point geometry, it would be advisable to replace the word 'sphere' by 'ball'.

Tarski worked with open balls. S. Jaśkowski — in the paper [Jaśkowski, 1948] cited by Tarski in [Tarski, 1929a] (see fn. 2, p. 26) and by Szczerba in [Szczerba, 1986, p. 911] — simplified Tarski's geometry by modifying the definition of the relation 'A is concentric with B', and working with closed instead of open balls.

the relation I above in terms of balls and the part-whole relation. The definitions are as follows:

Definition. A *point* is the class of all balls which are concentric with a given ball. (See Definition 6, p. 27).[130]

Definition. $I(x, y, z)$ iff there exists a ball W which belongs to the point x and which satisfies the following condition: no ball V belonging to the point y or the point z is a part of W or disjoint from W. (See Definition 7, p. 27).[131]

Given Pieri's result (above), it follows that every concept of 3-dimensional Euclidean geometry can be defined in terms of balls and the part-whole relation — and the result extends to n-dimensional Euclidean geometry for $n \geq 2$. As Szczcerba remarks (in [Szczerba, 1986]), the notion of a ball seems much more intuitive than the notion of a point — accordingly, this system of solid geometry is known as *Tarski's natural geometry*.

7.3 The Banach-Tarski paradox

The famous Banach-Tarski paradox is a startling theorem proved by Tarski and Stefan Banach in their paper "Sur la décomposition des ensembles de points en parties respectivement congruentes" [Tarski, 1924d]. This was Tarski's second published paper in geometry, and like the first — [Tarski, 1924b] — it is concerned with the notion of *equivalence by finite decomposition*. According to the classical definition of this notion, two geometric figures are equivalent by finite decomposition if one can be decomposed (or "cut up") into finitely many figures that can be rearranged to form the other one. Tarski worked with a generalized set-theoretical version of this classical notion, which treats geometric figures such as polygons and polyhedra as sets of points. Tarski characterizes *equivalence by finite decomposition* along these lines:

> Two sets of points A and B (in particular, two polygons or polyhedra) are said to be *equivalent by finite decomposition* iff A can be decomposed into finitely many disjoint sets of points $A_1, A_2, \ldots, A_n$, B can be decomposed into the same number of disjoint sets of points $B_1, B_2, \ldots, B_n$, and A_i and B_i are congruent, for $1 \leq i \leq n$.[132]

In [Tarski, 1924b] Tarski considers two theorems that were known to hold for the classical notion of equivalence by finite decomposition:

Theorem 1 Two arbitrary polygons, where one is contained in the other, are never equivalent by finite decomposition.

Theorem 2 Two polygons are equivalent by finite composition if and only if they

[130] Tarski observes (in [Tarski, 1929a, p. 27, fn. 1]) that balls are (first-order) individuals, while points are second-order objects. In point geometry, the reverse is true.

[131] Intuitively, the lines xy and xz are radii of the ball W.

[132] See Tarski [1924b, p. 64].

are equal in area.

Tarski shows in [Tarski, 1924b] that these theorems also hold for the set-theoretical version of equivalence by finite decomposition.

These results about polygons appear natural and intuitive, and one might expect them to hold for polyhedra too. But they do not. At the very end of [Tarski, 1924b], Tarski remarks that, if we take equivalence by finite decomposition in the set-theoretical sense, it can be shown that Theorems 1 and 2 do not extend to polyhedra. Tarski announces the following theorem, proved in the yet-to be-published paper [Tarski, 1924d]:

> Two arbitrary polyhedra (of the same volume or not) are equivalent by finite decomposition. (See [Tarski, 1924b, p. 64].)

As Tarski cautiously puts it, the result "seems perhaps paradoxical" (p. 64). This theorem is the Banach-Tarski paradox.[133]

The theorem is more precisely stated in [Tarski, 1924d] as follows:

> If A and B are any two sets of points in 3-dimensional Euclidean space, each being bounded and with a non-empty interior, then A and B are equivalent by finite decomposition.[134]

In particular, then, two spheres of different radii are equivalent by finite decomposition (in sharp contrast to the 2-dimensional case of circles). This encourages dramatic statements of the paradox:

> "It is possible to cut up a pea into finitely many pieces that can be rearranged to form a ball the size of the sun!"[135]

or

> "A ball, which has a definite volume, may be taken apart into finitely many pieces that may be rearranged via rotations of R^3 to form two, or even 1,000,000 balls, each identical to the original one."[136]

The theorem extends to n-dimensional Euclidean space, for $n \geq 3$.[137]

[133] Previous paradoxical decompositions were constructed by Vitali [Vitali, 1905] and Hausdorff [Hausdorff, 1914]. Hausdorff's work was a major influence on Banach and Tarski's paper. For more historical background, see [Wagon, 1985, Chapters 1–3].

[134] See [Tarski, 1924d, p. 140, Theorem 24]. In the statement of the theorem, Tarski uses the term "ensembles-frontières". This term is defined in fn. 1, p. 133: it applies to exactly those sets of points that have no interior point.

[135] [Wagon, 1985, p. xiii].

[136] [Wagon, 1985, p. 28].

[137] See [Tarski, 1924d, p. 121], where the theorem is stated in its fully general form, and also p. 137, where Tarski indicates how to proceed with the proof for $n \geq 3$. Moore writes that Banach and Tarski independently discovered the Banach–Tarski paradox, and agreed to write a joint paper. In a personal communication to Moore, Tarski reported that the general result was due to him (see [Moore, 1982, p. 285, fn. 3]).

Jan Mycielski calls the Banach-Tarski result "the most surprising result of theoretical mathematics".[138] One kind of response to the paradox — the most common — has been to take the result at face value, and accept it as a demonstration that mathematics can sometimes fail in the most radical way to match up with physical reality.[139] A different kind of response — little seen nowadays — has been to treat it as the conclusion of a genuine paradox, a conclusion that is so counterintuitive that we are forced us to review the argument that produced it. For some the culprit is the Axiom of Choice, which is indispensable to the proof.[140]

8 SET THEORY

Tarski's first published paper [Tarski, 1921] and his last published monograph [Tarski, 1987a] were both works in set theory. Tarski's early interest in set theory was inspired by Sierpiński, who, together with Janiszewski and Mazurkiewicz, started up the Warsaw School of Mathematics in 1919. In the 1920s and 1930s, Tarski wrote a number of papers on topics in general set theory. The subject of [Tarski, 1921] was the notion of a well-ordered set; subsequent papers dealt with such topics as the theory of finite sets, cardinal arithmetic, and the Axiom of Choice and its equivalents. As set theory developed and grew more sophisticated and specialized, so did Tarski's work. Tarski was a seminal figure in the development of the theory of large cardinals — he produced highly influential work on large cardinals from 1930 to the 1960s. Tarski also investigated set theory from the algebraic perspective, a perspective which — as we have seen — he so often brought to logic. In *Cardinal Algebras* [Tarski, 1949f] and *Ordinal Algebras* [Tarski, 1956c] Tarski investigated cardinal and ordinal addition within the framework of abstract algebraic systems. And in *A Formalization of Set Theory without variables* [Tarski, 1987a], Tarski and Steven Givant developed set theory within the framework of abstract relation algebras (see section 6.2 above for more on relation algebras).

8.1 General set theory

In his paper [Sierpiński, 1918], Sierpiński investigated in detail the role that the Axiom of Choice played in set theory and analysis, and threw out a challenge to

[138] See Mycielski's Foreword to [Wagon, 1985].

[139] The mathematical lesson to learn is that the Banach–Tarski theorem

> "precludes the existence of finitely-additive, congruent-invariant measures over all bounded subsets in [three-dimensional space] R^3."

(Mycielski, in the Foreword to [Wagon, 1985])

[140] For a detailed discussion of the role of the Axiom of Choice in the proof of the Banach-Tarski theorem, see [Wagon, 1985, Chapter 13]. Wagon's book is a recent mathematical work motivated by the Banach–Tarski result. It contains a number of open problems, old and new. One of these Wagon calls "Tarski's Circle-Squaring Problem" (posed in 1925): Is a circle (with interior points) equivalent by finite decomposition to a square of the same area? (See [Wagon, 1985, p. 101ff] for a discussion of this and related problems.)

mathematicians: determine the deductive relations between the Axiom of Choice and other propositions.[141] Tarski responded to this challenge in [Tarski, 1924a]. At the outset of this paper, Tarski lists seven propositions that are equivalent to the Axiom of Choice. Each of these propositions is a statement of cardinal arithmetic, where m, n, p, q are infinite cardinal numbers:

I. $m \cdot n = m + n$

II. $m = m^2$

III. If $m^2 = n^2$ then $m = n$

IV. If $m < n$ and $p < q$ then $m + p < n + q$

V. If $m < n$ and $p < q$ then $m.p < n.q$

VI. If $m + p < n + p$ then $m < n$

VII. If $m \cdot p < n \cdot p$ then $m < n$[142]

In their joint paper [Tarski, 1926a], Tarski and Lindenbaum added to this list seven more propositions, all of them statements from the theory of cardinal numbers (see [Tarski, 1926a, pp. 185–6, Theorem 82]).[143] Thus, Tarski's early work in [Tarski, 1924a] and [Tarski, 1926a] was a major contribution to our understanding of the relation between cardinal arithmetic and the Axiom of Choice, and in particular to the project of finding so-called 'cardinal equivalents' of Choice.[144]

Another topic of Tarski's early work in set theory was the theory of finite sets. In [Tarski, 1924c] Tarski systematically constructed a theory of finite sets on the basis of Zermelo's first five axioms (Extensionality, Elementary sets, Separation, Power set, Union) — Zermelo's remaining two axioms, the Axiom of Choice and the Axiom of Infinity, were excluded. Tarski remarks that no previous attempt to construct such a theory had been completely successful; for example, Dedekind's treatment in [Dedekind, 1888] did not have a solid axiomatic foundation, and Russell and Whitehead's account (in [Russell and Whitehead, 1910–13]) is specific to their theory of types (see [Tarski, 1924c, pp. 67–8].)

Tarski's own starting point was a new definition of finite set:

[141] See [Moore, 1982, Chapter 4.1] for more on Sierpiński' work and influence.

[142] Tarski's proof of the equivalences is based on Zermelo's system of axioms (omitting Choice, of course) together with two additional axioms which introduce the notion of cardinal number:
1. Every set has a cardinal number.
2. Two sets have the same cardinal number if and only if there is a 1-1 correspondence between them.

[143] Tarski and Lindenbaum's [Tarski, 1926a] was an extensive summary of their recent results in set theory, presented without proofs.

[144] In [Rubin and Rubin, 1963], Herman and Jean Rubin present over one hundred equivalents of the Axiom of Choice — and one category of equivalents they call 'cardinal equivalents'. For more on Tarski's work on cardinal equivalents, see [Moore, 1982, pp. 213–219, Section 4.3]. Other work on cardinal arithmetic published by Tarski in the 1920s includes [Tarski, 1925; Tarski, 1926a; Tarski, 1929c]. Tarski later obtained further cardinal equivalents of the Axiom of Choice in [Tarski, 1938a], a paper that investigated inaccessible cardinals (see section 8.2 below).

> A set A is *finite* iff every non-empty set K of subsets of A has a $\subseteq$-least element B (that is, an element B such that no member of K is a proper subset of B). (See [Tarski, 1924c, p. 71, Definition 3].)

From this starting point, Tarski went on to prove a large number of theorems which established all the fundamental properties of finite sets with which he was familiar. Tarski's definition differs from the standard arithmetic definition, according to which a set is finite iff it is either empty or there exists a 1-1 correspondence between it and the set $\{1, 2, \ldots, n\}$ for some finite n. In contrast to this usual definition, Tarski's definition is independent of the concept of finite number. It also differs from Dedekind's, according to which a set is finite if there is no 1-1 correspondence between it and any of its proper subsets.

The question naturally arises as to the equivalence of these various definitions, and this question is also investigated in [Tarski, 1924c]. Tarski shows that his definition is equivalent to the usual arithmetical one, thereby showing that the theory of finite sets can be constructed from the usual definition without the need for the Axiom of Choice. He also shows if a set is finite in his sense, then it is finite in Dedekind's sense. For Tarski, the reverse direction was an open problem: without Choice, it was not known how to establish the equivalence of his definition and Dedekind's (see [Tarski, 1924c, p. 95]). In an appendix to [Tarski, 1924c], Tarski states a number of open problems of this kind, generated by five alternative definitions of finite set that Tarksi lists (on p. 115). It turned out that none of these definitions — among them Tarski's and Dedekind's — were equivalent in the absence of the Axiom of Choice. So while Tarski's definition permitted the construction of the theory of finite sets without Choice, a number of alternative definitions (including Dedekind's) did not.[145]

8.2 Large cardinals

Tarski initiated the systematic study of large cardinals, and he and his school were largely responsible for its continuing development. An appropriate starting point is Tarski and Sierpiński's joint paper [Tarski, 1930a] in which they define the notion of an *inaccessible cardinal*. Their definition runs as follows:

> A cardinal number m is inaccessible iff it is not the product of a fewer number of cardinals of lesser power. (See Definition 1, p. 289).

The intuitive idea is that an inaccessible cardinal cannot be obtained from below. As Tarski and Sierpiński remark, $\aleph_0$ is clearly an inaccessible cardinal: it cannot be obtained from the product of finitely many finite cardinals. In current terminology, Tarski and Sierpiński's definition characterizes the class of *strongly inaccessible cardinals*.

In [Tarski, 1938a], Tarski provided alternative characterizations of the strongly inaccessible cardinals. He also characterized the weakly inaccessible cardinals, as follows:

[145] For more on Tarski on finite sets, see [Moore, 1982, pp. 209–213, section 4.2].

> m is a *weakly inaccessible cardinal* iff $m = \aleph_\alpha$ where α is a regular limit ordinal.[146]

Tarski shows, with the help of the Axiom of Choice, that if a cardinal is strongly inaccessible, it is weakly inaccessible. He also proves the reverse direction, but, as he makes fully explicit, he does so only with the help of the Generalized Continuum Hypothesis.

Tarski took inaccessible cardinals to be natural objects of set theory:

> "... the belief in the existence of inaccessible cardinals $> \omega$ (and even of arbitrarily large cardinals of this kind) seems to be a natural consequence of basic intuitions underlying the "naïve" set theory and referring to what can be called "Cantor's absolute"." [Tarski, 1962, p. 124]

And Zermelo's set theory cannot accommodate these intuitions:

> "on the basis of the usual axioms of Zermelo the existence of such numbers, apart from the two smallest of them, 2 and $\aleph_0$, cannot be established at all." [Tarski, 1939, p. 557]

Accordingly, Tarski introduced for the first time an axiom that guaranteed the existence of large cardinals.[147] Remarking that the axiomatization of increasingly large segments of "Cantor's absolute" is regarded by many as one of main aims of research in the foundations of set theory, Tarski continues:

> "Those who share this attitude are always ready to accept new "construction principles", new axioms securing the existence of new classes of "large" cardinals (provided they appear to be consistent with old axioms)." [Tarski, 1939, p. 557]

Tarski's *Axiom of Inaccessible Cardinals* was first presented in [Tarski, 1938a, p. 375]. Tarski's formulation was based on a technical characterization of the strongly inaccessible cardinals (drawn from [Tarski, 1938a, Theorem 21]). In [Tarski, 1939, p. 558], Tarski replaced it by a more natural version:

> For every set N there exists a set M with the following properties:

[146] For the notion of a regular ordinal, we need some prior notions. We call two partially ordered sets X and Y *similar* (in symbols, $X \sim Y$) if there exists an order-preserving 1-1 correspondence between them. Next we need the notions of an initial ordinal and cofinality.
Definition α is an *initial ordinal* iff $\mathrm{Ord}\alpha$ & $\forall\beta < \alpha(\neg\alpha \sim \beta)$.
Definitions
α is *cofinal* with β iff $\alpha \leq \beta$ & $\exists f(f : \alpha \to \beta$ & $\beta = \cup\mathrm{range}(f))$.
The *cofinality* $\mathrm{cf}(\beta)$ of β is $\cap\{\alpha|\alpha$ is cofinal with $\beta\}$.
We are now in a position to define the notion of a regular ordinal:
Definition For an initial ordinal κ, κ is *regular* iff $\mathrm{cf}(\kappa) = \kappa$.

[147] Compare the introduction of the Axiom of Replacement to guarantee the existence of certain intuitive sets — for example, the denumerable set $\{Z_0, PZ_0, PPZ_0, \dots\}$, where Z_0 is the set of positive integers and P is the power set operation (see [Skolem, 1922], in [van Heijenoort, 1967, pp. 296–7]).

(i) N is equipollent to a subset of M.[148]

(ii) the set of subsets of M which are not equipollent to M (i.e. the set $\{x|x \subseteq M \ \& \ x < M\}$)[149] is equipollent to M.

(iii) there is no set P such that P's power set is equipollent to M.

Either version of the Axiom of Inaccessible Sets says, in effect, that for any set there is a larger set whose cardinal number is strongly inaccessible. The axiom "assures the existence of inaccessible numbers as large as we please" [Tarski, 1939, p. 558]. Moroever, the axiom has "great deductive power":

> "If it is included in Zermelo's or Zermelo-Fraenkel's axiom-system this axiom brings with it a great simplification and reduction of the system; and, be it noted, the axiom of choice then becomes a provable theorem." [Tarski, 1939, p. 557]

In a later series of papers [Tarski 1943; 1957a; 1958a; 1958b; 1961a; 1962; 1964], Tarski investigated a family of problems concerning inaccessible cardinals. These problems share the same structure. The form of each problem is this: to determine the class of all the (infinite) cardinals that have a given property P, where it is known that $\aleph_0$ does not have P, and that the accessible cardinals do. The problem reduces to this: which of the inaccessible cardinals, if any, have P?

For example, one problem examined in [Tarski, 1962] is the *compactness problem* for predicate logics with infinitely long formulas.[150] Let L_ω be ordinary predicate logic. For any regular cardinal α let L_α be the logical system that differs from L_ω in having α different variables, infinitary operations analogous to ordinary disjunction and conjunction, and new versions of universal and existential quantification adjusted to the infinitary setting.[151] The compactness problem for logics L_α is the problem of determining those cardinals α for which the following compactness theorem holds:

> If S is any set of sentences in L_α, and if every subset of S with power $< \alpha$ has a model, then S also has a model.

That is, a cardinal α has the property P in the present case if the compactness theorem holds for α. If so, α is a *compact cardinal*; otherwise it is *incompact*. Now ω is compact, while all the accessible cardinals are incompact. What about the inaccessible cardinals? Here the problem about inaccessible cardinals has a metamathematical or metalogical setting. Tarski's student Hanf proved that a very large class of inaccessible cardinals are incompact.[152] Consider an arrangement of all the inaccessible cardinals in a transfinite increasing sequence $\theta_0, \theta_1, \ldots, \theta_i$,

[148] Two sets are equipollent iff there is a 1-1 correspondence between them.
[149] $A < B$ iff there is a set C such that A is equipollent to C and $C \subset B$.
[150] These logics are briefly discussed in [Tarski, 1958a].
[151] For a full description see, for example, [Tarski, 1962, pp. 116–7].
[152] [Hanf, 1963–4].

... for ordinals $0, 1, \ldots, i, \ldots$. Hanf proved that every cardinal α of the form $\alpha = \theta_i$ with $0 < i < \alpha$ is incompact.

In [Tarski, 1962] Tarski used this metamathematical result of Hanf's to establish certain other properties of this large class of inaccessible cardinals. Tarski still regarded the metamathematical approach as powerful and intuitive, even for results in pure set theory.[153] Referring back to [Tarski, 1962], Tarski and Keisler later wrote:

> "The results we have mentioned concerning large classes of inaccessible cardinals were originally obtained with the essential help of metamathematical (model-theoretic) methods. These methods still provide the intuitively and deductively simplest approach of the topic in its full generality. In our opinion this circumstance provides new and significant evidence of the power of metamathematics as a tool in purely mathematical research, and at the same time does not detract in the least from the value of the results obtained." [Tarski, 1964, p. 130]

Nevertheless in [Tarski, 1964] Tarski and Keisler undertake a purely mathematical (and very extensive) treatment of the topics discussed in [Tarski, 1962], motivated to some extent "by some (irrational) inclination toward puritanism in methods" [Tarski, 1964, p. 130].[154]

At the end of [Tarski, 1962], Tarski observes that the problems he has treated remain open only for the smallest inaccessible cardinal that is not incompact, and beyond. If there is no such cardinal, the problems are fully disposed of. And Tarski reports that

> "we do not know any example of a cardinal $\alpha > \omega$ which would possess a "constructive characterization" (in some very general and rather loose sense of this term) and of which we could not prove that it is incompact ..." [Tarski, 1962, p. 119]

So should we add an axiom to the effect that there are no such cardinals? Only if we are willing to compromise the study of "Cantor's absolute", which Tarski certainly was not. So we *should* be prepared to accept large cardinal axioms, but *not* prepared to accept

> "any axioms precluding the existence of [large] cardinals — unless this is done on a strictly temporary basis, for the restricted purpose of facilitating the metamathematical discussion of some axiomatic systems of set theory." [Tarski, 1962, p. 124]

[153] Work of Tarski's that adopts the metamathematical approach to the present family of problems are [Tarski, 1958a] (with D. Scott), [Tarski, 1961a; Tarski, 1961b]. See also [Keisler, 1962].

[154] Tarski [1962] and [Tarski, 1964] inspired a great deal of research on *measurable cardinals* and *ultraproducts* (see for example [Chang and Keisler, 1973, Ch. 4.2]. For some historical notes, see [Chang and Keisler, 1973, pp. 520–1]).

8.3 The algebraization of set theory

We turn first to Tarski's book *Cardinal Algebras* [Tarski, 1949f], published in 1949 but conceived some twenty years earlier. It deals with cardinal arithmetic, in particular cardinal addition. Tarski distinguishes two types of results in cardinal arithmetic, those that rely on the Axiom of Choice, and those, more constructive in nature, that mostly do not. In [Tarski, 1949f], Tarski investigates results of the second type that concern cardinal addition. Tarski establishes that all of these results can be derived within the framework of so-called *cardinal algebras*, which satisfy just a small number of laws or postulates. Moreover, these laws apply not only to cardinal arithmetic, but also to various other mathematical systems. So cardinal algebras form a simple framework within which many results from many different systems are forthcoming.

A cardinal algebra is an ordered triple $\underline{A} = \langle A, +, \Sigma\rangle$ satisfying certain postulates. A is a set of arbitrary elements, $+$ is a 2-place operation, and Σ an operation on infinite sequences (intuitively, Σ is infinite addition). Postulates I and II are respectively the closure of A under $+$ and the closure of A under Σ. Postulate III is the associative postulate: $\sum_{i<\infty} a_i = a_0 + \sum_{i<\infty} a_{i+1}$. Postulate IV is the commutative-associative postulate: $\sum_{i<\infty}(a_i + b_i) = \sum_{i<\infty} a_i + \sum_{i<\infty} b_i$. Postulate V is the zero postulate: there is an element z in A such that $a + z = z + a = a$ for every $a \in A$. Postulates VI and VII are the ones characteristic of cardinal algebras. Postulate VI is the refinement postulate: if $a + b = \sum_{i<\infty} c_i$ then there are a_i and b_i such that $a = \sum_{i<\infty} a_i$, $b = \sum_{i<\infty} b_i$, and $c_n = a_n + b_n$. Postulate VII is the remainder (or infinite chain) postulate: if $a_n = b_n + a_{n+1}$, for $n = 0, 1, 2, \ldots$ then there is an element $c \in A$ such that $a_n = c + \Sigma_{i<\infty} b_{n+i}$ for $n = 0, 1, 2, \ldots$. (See [Tarski, 1949f, pp. 3–4].) It is straightforward to check that the cardinal numbers with the operations of binary and infinite addition form a cardinal algebra. The first part of [Tarski, 1949f] develops the arithmetic of cardinal algebras. (The second part extends these results to a wider class of algebraic systems, and examines general methods of constructing cardinal algebras; the third part investigates the connections between cardinal algebras and related algebraic systems.) Tarski remarks:

> "The idea of an algebraic treatment of [cardinal arithmetic] and certain aspects and implications of the algebraic development seem to be essentially new." (Preface to [Tarski, 1949f, p. xii])

Tarski's monograph *Ordinal Algebras* [Tarski, 1956c] does for ordinal addition what [Tarski, 1949a] did for cardinal addition. As Tarski puts it:

> "The method applied in this monograph in the development and presentation of the theory of ordinal addition is the abstract algebraic method which was applied for analogous purposes in an earlier work of the author, [Tarski, 1949f]." (Introduction to [Tarski, 1956c], p. 2.)

Tarski remarks that the algebras themselves show some analogies — we find closure, associative, and remainder postulates. The main difference, according to

Tarski, is that in ordinal algebras the operation + is not commutative; also, the refinement postulate is replaced by a much stronger statement, the directed refinement postulate. (For a full specification of ordinal algebras, and more on these analogies and differences, see [Tarski, 1956c, Chapter 1]). As in the case of cardinal addition, Tarski's novel algebraic approach to ordinal addition generated, from a very small base, a large number of results that are not limited to ordinal arithmetic.[155]

Tarski's final monograph *A Formalization of Set Theory without variables* (with Steven Givant) [Tarski, 1987a] is a wholesale algebraization of set theory. In the preface, Tarski and Givant announce:

> "In this work we shall show that set theory and number theory can be developed within the framework of a new, different, and very simple formalism, L^X." (p. xi)

The formalism L^X contains no variables, quantifiers, or sentential connectives. The vocabulary of L^X contains seven symbols: two atomic predicates, $\mathring{1}$ (denoting the identity relation between individuals) and E (denoting the membership relation); four operators, $+$ (Boolean addition), $^-$ (Boolean complement), $^\smile$ (conversion — see 4.2 above), and $\odot$ (relative product — called ';' in 4.2 above); and a second identity predicate $=$ (denoting the identity relation between relations). Compound predicates are formed from the atomic predicates $\mathring{1}$ and E using the operators. All mathematical statements in L^X are of the form '$A = B$', where A and B are arbitrary predicates.

The deductive apparatus of L^X is based upon ten logical axiom schemata (see [Tarski, 1987a, p. 46]):

(I) $A + B = B + A$

(II) $A + (B + C) = (A + B) + C$

(III) $(A^- + B)^- + (A^- + B^-)^- = A$

(IV) $A \odot (B \odot C) = (A \odot B) \odot C$

(V) $(A + B) \odot C = A \odot C + B \odot C$

(VI) $A \odot \mathring{1} = A$

(VII) $A^{\smile\smile} = A$

(VIII) $(A + B)^\smile = A^\smile + B^\smile$

(IX) $(A \odot B)^\smile = B^\smile \odot A^\smile$

[155] Both [Jonsson, 1986] and [Levy, 1988] note that Tarski's work on cardinal and ordinal algebras has been somewhat neglected. Jonsson suggests that Tarski's thoroughness may in part explain this, "for he presented the mathematical community with two highly polished and elaborate creations, but with relatively few open problems" [Jonsson, 1986, p. 885].

(X) $\quad A^{\smile} \odot (A \odot B)^{-} + B^{-} = B^{-}.$

These axioms are the analogues of the equational postulates for abstract relation algebras essentially given by Chin and Tarski in [Tarski, 1951a, p. 344]. There is just one rule of inference, the rule of replacing equals by equals, familiar from high school algebra. Tarski and Givant write that

> "L^X proves adequate for the formalization of practically all known systems of set theory, and hence for the development of all of classical mathematics." (Preface to [Tarski, 1987a, p. xii])

Thus set theory can be based entirely on a system which is free of variables and in which the only formulas are sentences which take the form of equations. (See [Tarski, 1987a, section 4.6] for the formalizability of systems of set theory in L^X.) The system L^X is a direct development of Tarski's work on the calculus of relations and relation algebras (see 6.2 above):

> "Roughly speaking, the formalism L^X that is the central focus of this work is obtained from Tarski's equational formalization of the calculus of relations by introducing the constant E and deleting all variables." (Preface to [Tarski, 1987a, p. xvii])

Givant reports that Tarski died shortly after the manuscript for [Tarski, 1987a] was completed.

9 CONCLUDING REMARKS

Tarski wrote twenty monographs and more than one hundred articles, and yet he published next to nothing about his philosophical views. Even in his work best-known to philosophers — his work on the concept of truth — Tarski's philosophical views are hard to pin down, as we have seen.

In a rare expression of his philosophical standpoint regarding the foundations of mathematics, Tarski once wrote:

> "...I may mention that my personal attitude towards this question agrees in principle with that which has found emphatic expression in the writings of S. Leśniewski and which I would call *intuitionistic formalism*." [Tarski, 1930d, p. 62]

But in a footnote added later to this paper, Tarski characteristically drew back from such an unequivocal philosophical statement ("This last statement expresses the views of the author at the time when this article was originally published and does not adequately reflect his present attitude" (*ibid.*)).

Mostowski (Tarski's first PhD student) reports that Tarski was influenced early on by Kotarbinski's "reism", a nominalist doctrine, and would express sympathy with nominalism in oral discussions. But as Mostowski points out, this seems to

conflict with Tarski's constant use of abstract and general notions (consider, for example, Tarski's wholesale acceptance of set theory).[156]

We can find some brief philosophical remarks in two other pieces of Tarski's, both published posthumously. In a letter to Morton White, written in September 1944, Tarski followed what he took to be the Millian view, that logical and mathematical truths are, like empirical truths, the results of accumulated experience. Tarski goes on to say:

> "I think I am ready to reject certain logical premisses (axioms) of our science in exactly the same circumstances in which I am ready to reject empirical premisses (e.g. physical hypotheses)..." [Tarski, 1987b, p. 31]

This was a view that was later to find expression in [White, 1950] and [Quine, 1951], both of whom acknowledge their debt to Tarski.

The second posthumous piece is the paper "What are logical notions?" [Tarski, 1986a]. Tarski makes it clear that he is not after any 'platonic' answer to the question :

> "...people speak of catching the proper, true meaning of a notion, something independent of actual usage, and independent of any normative proposals, something like the platonic idea behind the notion. This last approach is so foreign and strange to me that I shall simply ignore it, for I cannot say anything intelligent on such matters." [Tarski, 1986a, p. 145]

Instead, Tarski aims to capture a possible use of the term 'logical notion', one that he thinks is in agreement with at least one actual usage of the term. As with the notions of definability, truth, and logical consequence, Tarski seeks to capture common usage as far as consistency and rigor allow.[157]

It is clear that Tarski's most significant philosophical contributions flow from his logical and metamathematical work, and not from any overtly philosophical writings. This is no accident: Tarski held firmly to the view that logical and mathematical investigations should proceed unhampered by any particular philosophical perspective. For example, as Mostowski emphasizes, Tarski's unrestricted use of set theory gave him a mathematical reach that was beyond that available to the adherents to Hilbert's formalism or Brouwer's intuitionism. Tarski thought of himself first and foremost as a mathematician and a logician, and as a philosopher only in some secondary sense: "perhaps a philosopher of a sort" [Tarski, 1944, p. 693].

Tarski's philosophical contributions are to be found in his work on metamathematics, semantics and logic. In the course of more than sixty years of active research, Tarski articulated concepts and established theorems that have become

[156] [Mostowski, 1967, p. 81].

[157] Tarski's suggestion concerning logical notions is couched in geometric terms of invariance under transformations.

standard in modern logic; he ushered in new fields; he helped to construct the very framework of modern logic. And through generation after generation of his students, Tarski's influence is still felt. In their review of [Tarski, 1956d], Pogorzelski and Surma write:

> "...the thing that we find most striking is that there ...is hardly another scientist in the history of the exact sciences whose part in the construction of notions for a large domain of science was as powerful as the contribution of Tarski to the creation of conceptual apparatus for logic, metalogic, and even metamathematics. In fact, the conceptual structure of these disciplines is due to Tarski."[158]

Tarski offered an additional way to measure "the value of a man's work" [Tarski, 1944, p. 693]:

> "It seems to me that there is a special domain of very profound and strong human needs related to scientific research, which are similar in many ways to aesthetic and perhaps religious needs. And it also seems to me that the satisfaction of these needs should be considered an important task of research. Hence, I believe, the question of the value of any research cannot be adequately answered without taking into account the intellectual satisfaction which the results of that research bring to those who understand it and care for it. It may be unpopular and out-of-date to say — but I do not think that a scientific result which gives us a better understanding of the world and makes it more harmonious in our eyes should be held in lower esteem than, say, an invention which reduces the cost of paving roads, or improves household plumbing." [Tarski, 1944, p. 694]

ACKNOWLEDGEMENTS

I am very grateful to my colleague Michael Resnik for his valuable comments on an entire draft of this chapter, and to David Ripley for his outstanding editorial work. I have also benefited from some excellent secondary sources. In particular, I have been helped by the following: the survey/expository papers on Tarski's work published by the *Journal of Symbolic Logic*, comprising Vaught [1986] (on model theory), Jonsson [1986] (on general algebra), McNulty [1986] (on undecidable theories), Monk [1986] (on algebraic logic), Szczerba [1986] (on geometry), Levy [1988] (on set theory), Van den Dries [1988] (on Tarski's elimination theory for real closed fields), Doner and Hodges [1988] (on decidable theories), Blok and Pigozzi [1988] (on general metamathematics), Etchemendy [1988a] (on truth and logical consequence), and Suppes [1988] (on the philosophical implications of Tarski's work); Corcoran [1983a]; Moore [1982]; Mostowski [1967].

[158] This is the conclusion of [Pogorzelski and Surma, 1969].

BIBLIOGRAPHY

WORKS BY TARSKI

[Tarski, 1921] 'Przyczynek do aksjomatyki zbioru dobrze uporzadkowanego' ('A contribution to the axiomatics of well-ordered sets'), Przeglad Filozoficzny, vol.24 , pp. 85–94, 1921.

[Tarski, 1923] 'O wyrazie pierwotnym logistyki', Doctoral dissertation, published in Przeglad Filozoficzny, vol.26, pp. 68–89, 1923. An English translation appears in [Tarski, 1983a, pp. 1–23], under the title 'On the primitive term of logistic'.

[Tarski, 1924a] 'Sur quelques theorèmes qui equivalent a l'axiome du choix', Fundamenta Matematyczno-fizyczny, vol.2, pp. 47–60, 1924.

[Tarski, 1924b] 'O rownowaznosci wielokatow' ('On the equivalence of polygons'), Przeglad Matematyczno-fizyczny, vol. 2, pp. 47–60, 1924. (Page references are to [Tarski, 1986b, vol. 1].)

[Tarski, 1924c] 'Sur les ensembles finis', Fundamenta Mathematicae, vol. 6, pp. 45–95, 1924. (Page references are to [Tarski, 1986b, vol. 1])

[Tarski, 1924d] with S. Banach. 'Sur la decomposition des ensembles de points en parties respectivement congruentes', Fundamenta Mathematicae, vol. 6, pp. 244–277, 1924. (Page references are to [Tarski, 1986b, vol. 1].)

[Tarski, 1925] 'Quelques theorèmes sur les alephs', Fundamenta Mathematicae, vol. 7, pp. 1–14, 1925.

[Tarski, 1926a] with A. Lindenbaum. 'Communication sur les récherches de la théorie des ensembles', Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie, Classe III, vol. 19, pp. 299–330, 1926. (Page references are to [Tarski, 1986b, vol. 1].)

[Tarski, 1927a] 'Remarque concernant l'arithmetique des nombres cardinaux' (abstract), Annales de la Société Polonaise de Mathematique, vol. 5, p. 101, 1927.

[Tarski, 1927b] with A. Lindenbaum. 'Sur l'independance des notions primitives dans les systemes mathématiques', Abstract, Annales de la Société Polonaise de Mathématique, vol. 5, pp. 111–113, 1927.

[Tarski, 1929a] 'Les fondements de la geometrie des corps', Ksiega Pamiatkowa Pierwszego Polskiego Zjazdu Matematycznego (Supplement to Rocznik Polskiego Towarzystwa Matematycznego), Krakow, pp. 29–33, 1929. A translation, with additions, appears in [Tarski, 1983a, pp. 24–29] under the title 'Foundations of the geometry of solids'. (Page references are to [Tarski, 1983a].)

[Tarski, 1929b] 'Remarques sur les notions fondamentales de la méthodologie des mathématiques', Annales de la Société Polonaise de Mathématique, vol. 7, pp. 270–2, 1929.

[Tarski, 1929c] 'Geschichtliche Entwicklung und gegenwartiger Zustand der Gleichmachtigkeitstheorie und der Kardinalzhlarithmetik', Ksiega Pamiatkowa Pierwszego Polskiego Zjazdu Matematycznego (Supplement to Rocznik Polskiego Towarzystwa Matematycznego), Krakow, pp. 48–54, 1929.

[Tarski, 1930a] with W. Sierpiński. 'Sur une propriété caracteristique des nombres inaccessibles', Fundamenta Mathematicae vol. 15, pp. 292–300, 1930. (Page references are to [Tarski, 1986b], vol. 1.)

[Tarski, 1930b] 'Über einige fundamentale Begriffe der Metamathematik', Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie, vol.23, cl.iii, pp. 22–29, 1930. An English translation appears in [Tarski, 1983a, pp. 30–37], under the title 'On some fundamental concepts of metamathematics'. (Page references are to [Tarski, 1983a].)

[Tarski, 1930c] with J. Łukasiewicz. 'Untersuchungen über den Aussagenkalkül', Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie, vol.23, c.iii, pp. 30–50, 1930. An English translation appears in [Tarski, 1983a, pp. 38–59], under the title 'Investigations into the sentential calculus'. (Page references are to [Tarski, 1983a].)

[Tarski, 1930d] 'Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften. I', Monatshefte fur Mathematik und Physik, vol. 37, pp. 361–404, 1930. An English translation appears in [Tarski, 1983a, pp. 60–109], under the title 'Fundamental concepts of the methodology of the deductive sciences'. (Page references are to [Tarski, 1983a].)

[Tarski, 1931a] 'Sur les ensembles definissables de nombres réels. I.', Fundamenta Mathematicae, vol. 17, pp. 210–239, 1931. An English translation appears in [Tarski, 1983a, pp. 110–142], under the title 'On definable sets of real numbers'. (Page references are to [Tarski, 1983a].)

[Tarski, 1931b] with C. Kuratowski. 'Les operations logiques et les ensembles projectifs' (with C. Kuratowski), Fundamenta Mathematicae, vol.17, pp. 240–248, 1931. An English translation appears in [Tarski, 1983a, pp. 143–151], under the title 'Logical operations and projective sets'. (Page references are to [Tarski, 1983a].)

[Tarski, 1933a] 'Pojecie prawdy w jezykach nauk dedukcyjnych' ('The concept of truth in the languages of deductive sciences'), Travaux de la Société des Sciences et des Lettres de Varsovie, Classe III, Sciences Mathématiques et Physiques, no.34, Warsaw, 1933. A German translation of this was published under the title 'Der Wahrheitsbegriff in den formalisierten Sprachen', in Studia Philosophica, vol. 1, pp. 261–405, 1935. An English translation of the German appears under the title 'The concept of truth in formalized languages' in [Tarski, 1983a, pp. 152–278]. (Page references are to [Tarski, 1983a].)

[Tarski, 1933b] 'Einige Betrachtungen über die Begriffe ω-Widerspruchsfreiheit und der ω-Vollstandigkeit', Monatshefte für Mathematik und Physik, vol. 40, pp. 97–112, 1933. An English translation appears in [Tarski, 1983a, pp. 279–295], under the title 'Some observations on the concepts of ω-consistency and ω-completeness'. (Page references are to [Tarski, 1983a].)

[Tarski, 1934] 'Z badan metodologicznych nad definiowalnosca terminow' in Przeglad Filozoficzny, vol. 37, pp. 438–460, 1934. A German translation appears under the title 'Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe', in Erkenntnis, vol. 5, pp. 80–100, 1935. An English translation of the German, supplemented by some passages translated from the Polish original, appears under the title 'Some methodological investigations on the definability of concepts' in [Tarski, 1983a, pp. 296–319]. (Page references are to [Tarski, 1983a].)

[Tarski, 1934–35] with A. Lindenbaum. 'Über die Beschränktheit der Ausdrucksmittel deduktiver Theorien', in Ergebnisse eines mathematischen Kolloquiums, fascicule 7, pp. 15–22, 1934–35. An English translation of the German appears under the title 'On the limitations of the means of expression of deductive theories', in [Tarski, 1983a, pp. 384–392]. (Page references are to [Tarski, 1983a].)

[Tarski, 1935] 'Zur Grundlegung der Booleschen Algebra I', in Fundamenta Mathematicae, vol. 24, pp. 177–98, 1935. An English translation appears under the title 'On the foundations of Boolean algebra' in [Tarski, 1983a, pp. 320–341]. (Page references are to [Tarski, 1983a].)

[Tarski, 1935+1936] 'Grundzüge des Systemenkalküls, Erster Teil', Fundamenta Mathematicae, vol. 25, pp. 503–526, 1935, and 'Grundzüge des Systemenkalküls, Zweiter Teil', Fundamenta Mathematicae, vol. 26, pp. 283–301, 1936. The two parts are translated as one paper in [Tarski, 1983a, pp. 342–383], under the title 'Foundations of the Calculus of Systems'. (Page references are to [Tarski, 1983a].)

[Tarski, 1936a] 'O ugruntowaniu naukowej semantyki' ('The establishment of scientific semantics'), Przeglad Filozoficzny, vol. 39, pp. 50–57, 1936. A German translation was published under the title 'Grundlegung der wissenschaftlichen Semantik', in Actes du Congrès International de Philosophie Scientifique, vol. 3, Actualités Scientifiques et Industrielle, vol. 390, pp. 1–8, Hermann & Cie, Paris, 1936. An English translation of the German appears under the title 'The establishment of scientific semantics' in [Tarski, 1983a, pp. 401–408]. (Page references are to [Tarski, 1983a].)

[Tarski, 1936b] 'O pojciu wynikania logicznego', Przeglad Filozoficzny, vol. 39, pp. 58–68, 1936. A German translation of this was published under the title 'Uber den Begriff der logischen Folgerung', in Actes du Congrès International de Philosophie Scientifique, vol. 7, Actualités Scientifiques et Industrielle, vol.394, pp. 1–11, Hermann & Cie, Paris, 1936. An English translation of the German appears under the title 'On the concept of logical consequence' in [Tarski, 1983a, pp. 409–420]. (Page references are to [Tarski, 1983a].)

[Tarski, 1936c] with A. Lindenbaum. 'Über die Beschranktheit der Ausdrucksmittel deduktiver Theorien' (with A. Lindenbaum), Ergebnisse eines Mathematischen Kolloquiums, vol. 7, pp. 15–22, 1936. An English translation appears under the title 'On the limitations of the means of expression of deductive theories' in [Tarski, 1983a, pp. 384–392]. (Page references are to [Tarski, 1983a].)

[Tarski, 1936d] O logice matematycznej i metodzie dedukcyjnej (On mathematical logic and the deductive method), Bibljoteczka Matematyczna, vol. 3–5, Ksiaznica-Atlas, Lwow and Warsaw, 1936.
[Tarski, 1938a] 'Über unerreichbare Kardinalzahlen', Fundamenta Mathematicae, vol. 30, pp. 68–89, 1938. (Page references are to [Tarski, 1986b, vol. 2].)
[Tarski, 1938b] 'Der Aussagenkalkul und die Topologie', Fundamenta Mathematicae, vol. 31, pp. 103–34, 1938. An English translation appears in [Tarski, 1983a, pp. 421–454], under the title 'Sentential Calculus and Topology'. (Page references are to [Tarski, 1983a].)
[Tarski, 1939] 'On well-ordered subsets of any set', Fundamenta Mathematicae, vol. 32, pp. 176–183, 1939. (Page references are to [Tarski, 1986b, vol. 2].)
[Tarski, 1941a] Introduction to logic and to the methodology of deductive sciences, Oxford University Press, Oxford and New York, 1941.
[Tarski, 1941b] 'On the calculus of relations', Journal of Symbolic Logic, vol. 6, pp. 73–89, 1941.
[Tarski, 1943] with P. Erdős. 'On families of mutually exclusive sets', Annals of Mathematics, ser. 2, vol. 44, pp. 315–329, 1943.
[Tarski, 1944] 'The semantic conception of truth and the foundations of semantics', Philosophy and Phenomenological Research, vol. 4, pp. 341–375, 1944. (Page references are to [Tarski, 1986b, vol. 2].)
[Tarski, 1948a] A Decision Method for Elementary Algebra and Geometry (prepared for publication by J. C. C. McKinsey), US Air Force Project RAND, R-109, the RAND Corporation, Santa Monica, California, 1948. (A second, revised edition was published by the University of California Press, Berkeley and Los Angeles, California 1951.)
[Tarski, 1948b] 'A problem concerning the notion of definability', Journal of Symbolic Logic, 13, 79–104, 1948.
[Tarski, 1949a] 'Arithmetical classes and types of Boolean algebras' (abstract), Bulletin of the American Mathematical Society, vol. 55, p. 64, 1949.
[Tarski, 1949b] 'On essential undecidability', Journal of Symbolic Logic, vol. 14, pp. 75–76, 1949.
[Tarski, 1949c] 'Undecidability in the arithmetic of integers and in the theory of rings', Journal of Symbolic Logic, vol. 14, p. 76, 1949.
[Tarski, 1949d] 'Undecidability of group theory' (abstract), Journal of Symbolic Logic, vol. 14, pp. 76–77, 1949.
[Tarski, 1949e] 'Undecidability of the theories of lattices and projective geometries' (abstract), Journal of Symbolic Logic, vol. 14, pp. 77–78, 1949.
[Tarski, 1949f] Cardinal Algebras, Oxford University Press, Oxford and New York, 1949.
[Tarski, 1951a] with L. H. Chin. 'Distributive and modular laws in the arithmetic of relation algebras', University of California Publications in Mathematics, new series, vol. 1, no. 9, pp. 341–384, 1951.
[Tarski, 1951b] with B. Jonsson. 'Boolean algebras with operators. Part I', American Journal of Mathematics, vol. 73, pp. 891–939, 1951.
[Tarski, 1952a] with B. Jonsson. 'Boolean algebras with operators. Part II', American Journal of Mathematics, vol. 74, pp. 127–162, 1952.
[Tarski, 1952b] with W. Szmielew. 'Mutual interpretability of some essentially undecidable theories', Proceedings of the International Congress of Mathematicians, Cambridge, Mass, 1950, vol. 1, American Mathematical Society, Providence, R.I., p. 734, 1952.
[Tarski, 1953] with A. Mostowski and R. M. Robinson. Undecidable theories, North-Holland Publishing Company, Amsterdam, 1953.
[Tarski, 1954a] 'Contributions to the theory of models, I', Indagationes Mathematicae, vol. 16, pp. 572–581, 1954. (Page references are to [Tarski, 1986b, vol. 3].)
[Tarski, 1954b] 'Contributions to the theory of models, II', Indagationes Mathematicae, vol. 16, pp. 582–588, 1954.
[Tarski, 1955] 'Contributions to the theory of models, III', Indagationes Mathematicae, vol. 17, pp. 56–64, 1955.
[Tarski, 1956a] with E. W. Beth. 'Equilaterality as the only primitive notion of Euclidean geometry', Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings, Series A, vol. 59 (= Indagationes Mathematicae, vol. 18), pp. 462–467, 1956.
[Tarski, 1956b] 'A general theorem concerning primitive notions of Euclidean geometry', Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings, Series A, vol. 59 (= Indagationes Mathematicae, vol. 18), pp. 468–474, 1956.

[Tarski, 1956c] Ordinal Algebras, North-Holland Publishing Company, Amsterdam 1956.
[Tarski, 1956d] Logic, Semantics, Metamathematics. Papers from 1923 to 1938, Clarendon Press, Oxford, 1956. (Translated by J. H. Woodger.)
[Tarski, 1957a] with E. C. Smith. 'Higher degrees of distributivity and completeness in Boolean algebras', Transactions of the American Mathematical Society, vol. 84, pp. 230–257, 1957.
[Tarski, 1957b] with R. L. Vaught. 'Arithmetical extensions of relational systems', Compositio Mathematica, vol. 13, pp. 81–102, 1957. (Page references are to [Tarski, 1986b, vol. 3].)
[Tarski, 1958a] with D. S. Scott. 'The sentential calculus with infinitely long expressions', Colloquium Mathematicum, vol. 6, pp. 165–170, 1958.
[Tarski, 1958b] 'Remarks on predicate logic with infinitely long expressions' Colloquium Mathematicum, vol. 6, pp. 171–176, 1958.
[Tarski, 1958c] with T. E. Frane and D. S. Scott. 'Reduced products', Notices of the American Mathematical Society, vol. 5, pp. 673–674, 1958.
[Tarski, 1959] 'What is elementary geometry?', The axiomatic method, with special reference to geometry and physics, L. Henkin *et al.*, editors, pp. 15–29. North-Holland Publishing Company, Amsterdam, 1959.
[Tarski, 1961a] with P. Erdős. 'On some problems involving inaccessible cardinals', Essays on the foundations of mathematics, Y. Bar-Hillel *et al.*, editors, pp. 50–82. Magnus Press, Jerusalem, 1961.
[Tarski, 1961b] 'Representable Boolean algebras and infinitary logics' (abstract), Notices of the American Mathematical Society, vol. 8, pp. 154–155, 1961.
[Tarski, 1961c] 'A model-theoretic result concerning infinitary logics' (abstract), Notices of the American Mathematical Society, vol. 8, p. 260, 1961.
[Tarski, 1962] 'Some problems and results relevant to the foundations of set theory', Logic, methodology, and philosophy of science. Proceedings of the 1960 International Congress, E. Nagel *et al.*, editors, pp. 125–135. Stanford University Press, Stanford, California, 1962. (Page references are to [Tarski, 1986b, vol. 4].)
[Tarski, 1964] wtih H. J. Keisler. 'From accessible to inaccessible cardinals. Results holding for all accessible cardinal numbers and the problem of their extension to inaccessible ones', Fundamenta Mathematicae, vol. 53, pp. 225–308, 1964. (Page references are to [Tarski, 1986b, vol. 4].)
[Tarski, 1965a] with L. W. Szczerba. 'Metamathematical properties of some affine geometries', Logic, methodology, and philosophy of science. Proceedings of the 1964 International Congress, Y. Bar-Hillel, editor, pp. 166–178. North-Holland Publishing Company, Amsterdam, 1965.
[Tarski, 1965b] with J. W. Addison and L. Henkin, eds. The theory of models, North-Holland, Amsterdam, 1965.
[Tarski, 1967] The completeness of elementary algebra and geometry, Institut Blaise Pascal, Paris, 1967.
[Tarski, 1969] 'Truth and Proof', Scientific American, vol. 220, no. 6, pp. 63–77, 1969. (Page references are to [Tarski, 1986b, vol. 4].)
[Tarski, 1971] with L. Henkin and J. D. Monk. Cylindric Algebras, Part I, North-Holland Publishing Company, Amsterdam, 1971.
[Tarski, 1978] with J. Doner and A. Mostowski. 'The elementary theory of well-ordering — A metamathematical study', Logic Colloquium '77, A. Macintyre *et al.*, editors, pp. 1–54. North-Holland Publishing Company, Amsterdam.
[Tarski, 1979] with L. W. Szczerba. 'Metamathematical discussion of some affine geometries', Fundamenta Mathematicae, vol. 104, pp. 155–92, 1979.
[Tarski, 1981] with H. Andreka, L. Henkin, J. D. Monk and I. Nemeti. Cylindric set algebras, Lecture Notes in Mathematics, vol. 883, Springer-Verlag, Berlin and New York, 1981.
[Tarski, 1983a] Logic, semantics, metamathematics. Papers from 1923 to 1938 (translations by J. H. Woodger). Second, revised edition of [Tarski, 1956d], with editor's introduction and an analytic index (J. Corcoran, editor), Hackett Publishing Company, Indianapolis, Indiana, 1983.
[Tarski, 1983b] with W. Schwabhauser and W. Szmielew. Metamathematische Methoden in der Geometrie, Hochschultext, Springer-Verlag, Berlin and New York, 1983.
[Tarski, 1985] with L. Henkin and J. D. Monk. Cylindric Algebras, Part II, North-Holland Publishing Company, Amsterdam 1985.

[Tarski, 1986a] 'What are logical notions?' John Corcoran, editor, History and Philosophy of Logic, 7, pp. 143–154, 1986.
[Tarski, 1986b] Affred Tarski: Collected Papers, vols. 1–4, S. Givant and R. McKenzie, eds., Birkhäuser, Basel, Boston, Stuttgart, 1986.
[Tarski, 1987a] with S. Givant. A formalization of set theory without variables, Colloquium Publications, vol. 41, American Mathematical Society, Providence, Rhode Island, 1987.
[Tarski, 1987b] 'A philosophical letter of Alfred Tarski', Journal of Philosophy, pp. 28–32, 1987.

WORKS BY OTHER AUTHORS

[Ajdukiewicz, 1921] K. Ajdukiewicz. Z metodologji nauk dedukcyjnych (From the methodology of the deductive sciences), Lwow, 1921.
[Aristotle, 1941] Aristotle, Posterior Analytics, pp. 108–186 of [McKeon, 1941].
[Barwise and Etchemendy, 1987] J. Barwise and J. Etchemendy. The Liar, Oxford University Press, 1987.
[Beth, 1953] E. W. Beth. 'On Padoa's method in the theory of definition', Indagationes Mathematicae, xv, pp. 330–9, 1953.
[Birkhoff, 1948] G. Birkhoff. Lattice Theory, American Mathematical Society Colloquium publications, vol. 25, revised edition, New York, 1948.
[Blackburn and Simmons, 1999] S. Blackburn and K. Simmons. Truth, in the series Oxford Readings in Philosophy, Oxford University Press, 1999.
[Blok and Pigozzi, 1988] W. J. Blok and D. Pigozzi. 'Alfred Tarski's work on general metamathematics', Journal of Symbolic Logic, vol. 53, no. 1, pp. 36–50, 1988.
[Bolzano, 1837] B. Bolzano. Wissenschaftslehre (Theory of Science), Sulzbach: Seidel, 4 vols., 1837.
[Burge, 1979] T. Burge. 'Semantical Paradox', Journal of Philosophy 76, pp. 169–98, 1979; reprinted in [Martin, 1984, pp. 83–117]. Page references are to [Martin, 1984].
[Carnap, 1934] R. Carnap. Logische Syntax der Sprache, Wien, 1934.
[Carnap, 1935] R. Carnap. 'Ein Gultigkeitskriterium für die Satze der klassischen Mathematik', Monatshefte für Mathematik und Physik, xlii, pp. 163–90, 1935.
[Chang and Keisler, 1973] C. C. Chang and H. J. Keisler. Model Theory, North-Holland Publishing Company, Amsterdam, 1973.
[David, 1994] M. David. Correspondence and Disquotation, Oxford University Press, 1994.
[Davidson, 1967] D. Davidson. 'Truth and Meaning', Synthese 17, pp. 304–23, 1967.
[Davidson, 1996] D. Davidson. 'On the Folly of Trying to Define Truth', Journal of Philosophy 93, pp. 263–278, 1996.
[Dedekind, 1888] R. Dedekind. Was sind und was sollen die Zahlen? Braunschweig: Vieweg, 1888.
[Doner and Hodges, 1988] J. Doner and W. Hodges. 'Alfred Tarski and decidable theories', Journal of Symbolic Logic, vol. 53, no. 1, pp. 20–35, 1988.
[Edwards, 1967] P. Edwards, ed. The Encyclopedia of Philosophy, Macmillan Publishing Co., New York, 1967.
[Etchemendy, 1983] J. Etchemendy. 'The doctrine of logic as form', Linguistics and Philosophy, vol. 6, pp. 319–334, 1983.
[Etchemendy, 1988a] J. Etchemendy. 'Tarski on truth and logical consequence', Journal of Symbolic Logic, vol. 53, no. 1, pp. 51–79, 1988.
[Etchemendy, 1988b] J. Etchemendy. 'Models, semantics and logical truth', Linguistics and Philosophy, vol. 11, pp. 91–106, 1988.
[Etchemendy, 1990] J. Etchemendy. The Concept of Logical Consequence, Harvard University Press, Cambridge 1990.
[Field, 1972] H. Field. 'Tarski's theory of truth', Journal of Philosophy, vol. LXIX, no. 13, pp. 347–375, 1972.
[Field, 1986] H. Field. 'The Deflationary Conception of Truth', in Fact, Science, and Morality: Essays on A. J. Ayer's 'Language, Truth and Logic', edited by G. MacDonald and C. Wright, pp. 55–117. Oxford: Basil Blackwell, 1986.
[Field, 1994] H. Field. 'Deflationist views of meaning and content', Mind 103, pp. 249–84, 1994.

[Gaifman, 1992] H. Gaifman. 'Pointers to Truth', Journal of Philsophy 89, pp. 223–61, 1992.
[Givant, 1986] S. Givant. 'Bibliography of Alfred Tarski', Journal of Symbolic Logic, vol. 51, no. 4, pp. 913–941, 1986.
[Glanzberg, 2001] M. Glanzberg. 'The Liar in context', Philosophical Studies 103, pp. 217–251, 2001.
[Gödel, 1930a] K. Gödel. 'Die Vollstandigkeit der Axiome des logischen Funktionenkalküls', Monatshefte fur Mathematik und Physik, 37, pp. 349–360; translated in [van Heijenoort, 1967, pp. 582–591].
[Gödel, 1930b] K. Gödel. 'Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit', Akademie der Wissenschaften in Wien, Mathematisch-naturwissenschaftliche Klasse, Akademischer Anzeiger, lxvii, pp. 214–215, 1930.
[Gödel, 1931] K. Gödel. 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I', Monatshefte für Mathematik und Physik, xxxviii, pp. 173–98, 1931.
[Gómez-Torrente, 1996] M. Gómez-Torrente. 'Tarski on logical consequence', Notre Dame Journal of Formal Logic, vol. 37, no. 1, pp. 125–151, 1996.
[Grzegorczyk, 1951] A. Grzegorczyk. 'Undecidability of some topological theories', Fundamenta Mathematicae, vol. 38, pp. 137–152, 1951.
[Hanf, 1963–4] W. P. Hanf. 'Incompactness in languages with infinitely long expressions', Fundamenta Mathematicae, vol. 53, pp. 309–324, 1963–4.
[Hausdorff, 1914] F. Hausdorff. 'Bemerkung über den Inhalt von Punktmengen' Mathematische Annalen 75, pp. 428–433, 1914.
[Herbrand, 1928] J. Herbrand. 'Sur la théorie de la demonstration', Comptes Rendus Hebdomadaires des Séances de l'Academie des Sciences, vol. 186, pp. 1274–1276, 1928.
[Herzberger, 1981] H. Herzberger. 'New Paradoxes for Old', Proceedings of the Aristotelian Society 81, pp. 109–23, 1981.
[Hilbert, 1922] D. Hilbert. Grundlagen der Geometrie, 3rd ed., Teubner, Leipzig, 1922.
[Hilbert and Bernays, 1939] D. Hilbert and P. Bernays. Grundlagen der Arithmetik, vol. 1, Berlin, 1934; vol.2, Berlin, 1939.
[Hodges, 1986] W. Hodges. 'Alfred Tarski', Journal of Symbolic Logic, vol. 51, no. 4, p. 866, 1986.
[Hook, 1950] S. Hook, ed. John Dewey: Philosopher of Science and Freedom, New York, Dial Press, 1950.
[Huntington, 1904] E. V. Huntington. 'Sets of independent postulates for the algebra of logic', Transactions of the American Mathematical Society, v, pp. 288–309, 1904.
[Jaśkowski, 1948] S. Jaśkowski. 'Une modification des definitions fondamentales de la geometrie des corps de M. A.Tarski', Annales de la Société Polonaise de Mathématique, xxi, pp. 298–301, 1948.
[Jonsson, 1986] B. Jonsson. 'The contributions of Alfred Tarski to general algebra', The Journal of Symbolic Logic, vol. 51, no. 4, pp. 883–889, 1986.
[Keisler, 1962] H. J. Keisler. ''Some applications of the theory of models to set theory', Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress, pp. 80–86. Stanford, 1962.
[Koons, 1992] R. C. Koons. Paradoxes of Belief and Strategic Rationality, Cambridge University Press, Cambridge, 1992.
[König, 1905] J. König. 'On the foundations of set theory and the continuum problem', in [van Heijenoort, 1967, pp. 145–149].
[Langford, 1927] C. H. Langford. 'Some theorems on deducibility', Annals of Mathematics, ser. 2, vol. 28, pp. 16–40, 1927.
[Langford, 1928] C. H. Langford. 'Theorems on deducibility (second paper)', Annals of Mathematics, ser. 2, vol. 28, 459–471, 1928.
[Leeds, 1978] S. Leeds. 'Theories of Reference and Truth', Erkenntnis 13, pp. 111–129, 1978.
[Leśniewski, 1916] S. Leśniewski. 'Podstawy ogolnej teorii mnogosci I' (Foundations of general set theory I) Moscow, 1916.
[Leśniewski, 1929] S. Leśniewski. 'Grundzuge eines neuen Systems der Grundlagen der Mathematik', Secs 1–11, Fundamenta Mathematicae, xiv, pp. 1–81, 1929.
[Leśniewski, 1927–31] S. Leśniewski. 'O podstawach matematyki' (On the foundations of mathematics), Przeglad Filozoficzny, xxx, pp. 164–206, 1927; ibid., xxxi, pp. 261–91, 1928; ibid., xxxii, pp. 60–105, 1929; ibid., xxxiii, pp. 77–105, 1930; ibid., xxxiv, pp. 142–70, 1931.

[Levy, 1988] A. Levy. 'Alfred Tarski's work in set theory', Journal of Symbolic Logic, vol. 53, no. 1, pp. 2–6, 1988.
[Löwenheim, 1915] L. Löwenheim. 'Über Moglichkeiten im Relativkalkül', Mathematische Annalen, vol. 76, pp. 447–470, 1915.
[Lyndon, 1950] R. Lyndon. 'The representation of relational algebras', Annals of Mathematics, ser. 2, vol. 51, pp. 707–729, 1950.
[Lyndon, 1956] R. Lyndon. 'The representation of relational algebras. II', Annals of Mathematics, ser. 2, vol. 63, pp. 294–307, 1956.
[Mal'cev, 1936] A. Mal'cev. 'Untersuchungen aus dem Gebiete der mathematischen Logik', Matematiceskii Sbornik, vol. 1 (43), pp. 323–336, 1936.
[Martin, 1976] R. L. Martin. 'Are Natural Languages Universal?', Synthese 32, pp. 271–91, 1976.
[Martin, 1984] R. L. Martin, ed. Recent Essays on Truth and the Liar Paradox, Oxford University Press, 1984.
[McKeon, 1941] R. McKeon, ed. The Basic Works of Aristotle, Random House, New York, 1941.
[McNulty, 1986] G. McNulty. 'Alfred Tarski and undecidable theories', The Journal of Symbolic Logic, vol. 51, no. 4, pp. 890–898, 1986.
[Moore, 1982] G. Moore, Zermelo's Axiom of Choice: its origins, development and influence, Springer Verlag, New York, 1982.
[Monk, 1986] J. D. Monk. 'The Contributions of Alfred Tarski to Algebraic Logic', The Journal of Symbolic Logic, vol. 51, no. 4, pp. 899–906, 1986.
[Mostowski, 1967] A. Mostowski. 'Alfred Tarski', in [Edwards, 1967, vol. 8, pp. 77–81].
[Padoa, 1900] A. Padoa. 'Logical introduction to any deductive theory', 1900. Translated in [van Heijenoort, 1967, pp. 118–123].
[Parsons, 1974] C. Parsons. 'The Liar Paradox', Journal of Philosophical Logic 3, pp. 381–412, 1974. Reprinted in [Martin, 1984]. Page references are to Martin 1984.
[Pasch, 1982] M. Pasch. Vorlesungen über neuere Geometrie, Leipzig: Teubner, 1882.
[Peano, 1895–1908] G. Peano. Formulaire de mathématiques, 5 vols. Turin: Bocca, pp. 1895-1908.
[Pieri, 1908] M. Pieri. 'La geometria elementare istituita sulle nozioni di 'punto' e 'sfera' '. Memorie di Matematica e di Fisica della Societa Italiana delle Scienze, ser. 3, 15, pp. 345–450, 1908.
[Pogorzelski and Surma, 1969] W. Pogorzelski and S. Surma. Review of [Tarski, 1956], Journal of Symbolic Logic 34, pp. 99–106, 1969.
[Presburger, 1930] M. Presburger. 'Über die Vollstandigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt', Comptes-rendus du I congrès des mathematiciens des pays slaves, Warszawa, 1929, Ksiaznica Atlas T.N.S.W. Warsaw, pp. 92–101, 1930.
[Quine, 1951] W. V. O. Quine. 'Two Dogmas of Empiricism', Philosophical Review, LX, 1, pp. 20–43, 1951.
[Quine, 1970] W. V. O. Quine. Philosophy of Logic, Englewood Cliffs, N.J.: Prentice-Hall, 1970.
[Resnik, 1990] M. Resnik. 'Immanent Truth', Mind, vol. 99, no. 395, pp. 405–424, 1990.
[Richard, 1905] J. Richard. 'Les principes des mathématiques et le problème des ensembles', in Revue générale des sciences pures et appliquées 16, p. 541, 1905. Also in Acta Mathematica 30, pp. 295–6, 1906. An English translation appears in [van Heijenoort, 1967, pp. 143–44].
[Robinson, 1949] J. Robinson. 'Definability and decision problems in arithmetic', Journal of Symbolic Logic, vol. 14, pp. 98–114, 1949.
[Robinson, 1950] R. M. Robinson. 'An essentially undecidable axiom system', Proceedings of the International Congress of Mathematicians, vol. 1, pp. 729–730, 1950.
[Robinson, 1951] R. M. Robinson. 'Undecidable rings', Transactions of the American Mathematical Society, vol. 70, pp. 137–159, 1951.
[Rubin and Rubin, 1963] H. Rubin and J. Rubin. Equivalents of the Axiom of Choice, Amsterdam, North-Holland, 1963.
[Russell, 1908] B. Russell. 'Mathematical logic as based on the theory of types', American Journal of Mathematics 30, pp. 222–262, 1908. Reprinted in [van Heijenoort, 1967, pp. 150–182].
[Russell and Whitehead, 1910–13] B. Russell and A. N. Whitehead. Principia Mathematica, Cambridge, Cambridge University Press: vol.1, 1910; vol. 2, 1912; vol. 3, 1913.
[Schröder, 1890–1905] E. Schröder. Vorlesungen über die Algebra der Logik (exakte Logik), Leipzig; i, 1890; ii, part 1, 1891; ii, part 2, 1905; iii, part 1, 1895.

[Schröder, 1895] E. Schröder. Algebra und Logik der Relative, 1895, vol.iii, part 1 of [Schröder, 1890–1905].

[Sierpiński, 1918] W. Sierpiński. 'L'axiome de M. Zermelo et son role dans la théorie des ensembles et l'analyse', Bulletin de l'Academie des Sciences de Cracovie, Classe des Sciences Math., Serie A, pp. 97–152, 1918.

[Simmons, 1993] K. Simmons. Universality and the Liar, Cambridge University Press, Cambridge and New York, 1993.

[Skolem, 1919] T. Skolem. 'Untersuchungen über die Axiome des Klassenkalküls und über Produktations — und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen', Skrifter utgit av Videnskapsselskapet i Kristiania, I. klasse 1919, no. 3, Oslo, 1919.

[Skolem, 1920] T. Skolem. 'Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen', Videnskapssel Skrifter I. Matematisk-naturvidenskabelig Klasse, vol. 6, pp. 1–36, 1920. Translated under the title 'Logico-combinatorial investigations in the satisfiability or provability of mathematical propositions: A simplified proof of a theorem by L. Löwenheim and generalization of the theorem', in [van Heijenoort, 1967, pp. 254–63].

[Skolem, 1922] T. Skolem. 'Some remarks on axiomatized set theory', 1922. Reprinted in [van Heijenoort, 1967, pp. 290–301].

[Skolem, 1934] T. Skolem. 'Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen, Fundamenta Mathematicae, vol.23, pp. 150–161, 1934.

[Soames, 1999] S. Soames. Understanding Truth, Oxford University Press, Oxford and New York, 1999.

[Stone, 1936] M. H. Stone. 'The theory of representation for Boolean algebras', Transactions of the American Mathematical Society, xi, pp. 37–111, 1936.

[Stone, 1937] M. H. Stone. 'Applications of the theory of Boolean rings to general topology', Fundamenta Mathematicae, xxix, pp. 223–303, 1937.

[Stone, 1937–38] M. H. Stone. 'Topological representations of distributive lattices and Brouwerian logics', Casopis pro Pestovani Matematiky a Fysiky, lxvii, pp. 1–25, 1937–8.

[Suppes, 1988] P. Suppes. 'Philosophical implications of Tarski's work', Journal of Symbolic Logic, vol. 53, no. 1, 80–91, 1988.

[Szczerba, 1986] L. W. Szczerba. 'Tarski and geometry', The Journal of Symbolic Logic, vol. 51, no. 4, 907–912, 1986.

[van den Dries, 1988] L. van den Dries. 'Alfred Tarski's elimination theory for real closed fields', Journal of Symbolic Logic, vol. 53, no. 1, 7–19, 1988.

[van Heijenoort, 1967] J. van Heijenoort, ed. From Frege to Gödel: a source book in mathematical logic, Harvard University Press, 1967.

[Vaught, 1965] R. L. Vaught. 'A Löwenheim-Skolem theorem for cardinals far apart', Proceedings of the 1963 Berkeley symposium on the theory of models, J. Addison *et al.*, editors, pp. 390–401. North-Holland, Amsterdam, 1965.

[Vaught, 1974] R. L. Vaught. 'Model theory before 1945', Proceedings of the Tarski symposium, Proceedings of Symposia in Pure Mathematics, vol. 25, pp. 153–172. American Mathematical Society, Providence, Rhode Island, 1974.

[Vaught, 1986] R. L. Vaught. 'Alfred Tarski's work in model theory', The Journal of Symbolic Logic, vol. 51, no. 4, 869–882, 1986.

[Veblen and Young, 1910] O. Veblen and J. W. Young. Projective Geometry, Ginn, Boston, Massachusetts (2 vols., first, 1910; second, 1918); reprint, Blaisdell, New York, 1938.

[Vitali, 1905] G. Vitali. Sul problema della mesura dei gruppi di punti di una retta, Bologna, Tip. Gamberini e Parmeggiani, 1905.

[Wagon, 1985] S. Wagon. The Banach-Tarski Paradox, Cambridge University Press, 1985.

[White, 1950] M. White. 'The Analytic and the Synthetic: An Untenable Dualism'. In [Hook, 1950, pp. 316–30].

EMIL POST

Alasdair Urquhart

1 BIOGRAPHY

Emil Leon Post was born on February 11 1897 into an Orthodox Jewish family in Augustów, a town at that time within the Russian empire, but after 1918 in the province of Białystok in Eastern Poland. In 1897 his father Arnold emigrated to join his brother in America. Seven years later, in May 1904, with the success of the family clothing and fur business in New York, his wife Pearl, together with Emil and his sisters Anna and Ethel, joined him. The family lived in a comfortable home in Harlem.



Figure 1. Emil Post, June 1924

As a child, Post was particularly interested in astronomy, but an accident at the age of twelve foreclosed that choice of career. As he reached for a lost ball under a parked car, a second car crashed into it, and he lost his left arm below the shoulder. As a high school senior, Post wrote to several observatories inquiring whether his handicap would prevent him pursuing a career as an astronomer. The responses he received, though not uniformly negative, were sufficient to discourage him from following his childhood ambition; instead, he turned towards mathematics.

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

Post attended Townsend Harris High School, a free secondary school for gifted students on the campus of City College, and received a B.S. in mathematics from City College in 1917 (where he benefited from a college education free of tuition fees). While still an undergraduate, he had already done a good deal of the work for a paper on generalized differentiation, presented to the October 1923 meeting of the American Mathematical Society, but only published in full seven years later [Post, 1930]. Post was a graduate student at Columbia University from 1917 to 1920, and it was at Columbia that his interest in modern mathematical logic was aroused by a seminar on the recently published *Principia Mathematica* [Whitehead and Russell, 1910-13] conducted by Cassius J. Keyser. His doctoral thesis, written under the supervision of Keyser, is a systematic study of the propositional calculus of *Principia*, including a proof of completeness.

Post spent the academic year 1920-21 as a Procter Fellow at Princeton University, and it was during this postdoctoral year that he discovered results anticipating the later incompleteness theorems of Kurt Gödel and the undecidability results of Church and Turing. The excitement caused by these startling discoveries precipitated his first attack of manic-depressive illness, a condition that plagued him for the rest of his life. He recovered sufficiently from this first attack to take up an instructorship at Cornell University, but a second attack led to a withdrawal from university teaching, and during the 1920s, he supported himself by teaching at George Washington High School in New York.

Under the care of a general practitioner, Dr. Levy, Post developed a routine that was designed to avoid undue excitement leading to manic attacks. His regimen involved strictly restricting the time spent on his research to three hours a day, from 4 to 5 p.m. and then from 7 to 9 p.m.

In 1932, Post was appointed to the faculty of City College of New York. He left after only one month, but returned in 1935 for the rest of his career. In spite of restrictions on his research time due to his treatment regimen and a teaching load of sixteen hours a week, Post was able to publish some of his most remarkable and influential papers during this period. His marriage to Gertrude Singer in 1929 undoubtedly introduced a measure of stability into his life, and she assisted him both by typing his papers and letters, and taking care of day to day financial affairs. Their daughter Phyllis Goodman remarks that: "My mother ... was the buffer in daily life that permitted my father to devote his attention to mathematics (as well as to his varied interests in contemporary world affairs). Would he have accomplished so much without her? I, for one, don't think so" [Davis, 1994, p. xii].

Post was a remarkably successful and popular teacher at City College; his unusual pedagogical methods are described below in §8. He continued to struggle with manic-depressive illness, and in 1954, he succumbed once more. He died of a heart attack on 21 April 1954 after electro-shock treatment at an upstate New York mental hospital.

2 INFLUENCES ON POST

To a remarkable extent, Post as a logician was a homegrown American original. Most of the other outstanding logicians of the 1920s and 1930s, such as Gödel, Herbrand, Bernays, Turing and Church, were influenced more or less directly by Hilbert's formalist program, and the problems that grew out of it. Post, on the other hand, after the initial stimulus provided by *Principia Mathematica*, created his own research project and problems more or less single handed, at a time when logic and the foundations of mathematics were not considered respectable topics for mathematical work in the United States.

An important early influence on his ideas about logic was the philosopher-logician C.I. Lewis. Lewis's monograph *A Survey of Symbolic Logic* [Lewis, 1918] is an expository work largely devoted to bridging the gap between the tradition of algebraic logic of Boole and Schröder and *Principia Mathematica*, but it has a remarkable sixth chapter that had a strong effect on Post's early thinking. Lewis contrasts the logistic method of Whitehead and Russell with what he calls a "heterodox" view of the nature of mathematics and logistic.

Lewis points out that the foundational portion of *Principia Mathematica* falls short of being a completely formal system. Rather, the primitive propositions of that work are presented in terms of the intuitive meanings associated with the notation. Thus, for example, the formation rules of the language are presented by phrases such as "If p is an undetermined elementary proposition, "not-p" is an elementary propositional function," while the rule of detachment (*modus ponens*) is stated as follows: "Anything implied by a true elementary proposition is true" [Whitehead and Russell, 1910-13, Vol. 1, pp. 92, 94]. Whitehead and Russell fail to make the basic distinction between axioms and rules of inference, since they are lumped together under the heading of "Primitive Propositions."

Lewis presents his heterodox view of the logistic method in the following striking definition:

> *A mathematical system is any set of strings of recognizable marks in which some of the strings are taken initially and the remainder derived from these by operations performed according to rules which are independent of any meaning assigned to the marks* [Lewis, 1918, p. 355].

Lewis illustrates his definition by presenting the propositional calculus of *Principia Mathematica* in a strictly formalist style. He uses the purposely meaningless words "quid" and "quod" to refer to formulas and propositional connectives. With this terminology, he restates the formation rule above as follows: "The combination of any quid preceded immediately by the mark $\sim$ may be treated as a quid." The rules of substitution and *modus ponens* he states in this way:

> (7) In any string in the initial set, or in any string added to the list according to rule, any quid whatever may be substituted for p or q or r, or for any quid consisting of only one mark. When a quid is substituted

> for any mark in a string, the same quid must also be substituted for that same mark wherever it appears in the string.
>
> (8) The string resulting from the substitution of a quid consisting of more than one mark for a quid of one mark, according to (7), may be added to the list of strings.
>
> (9) In any string added to the list, according to (8), if that portion of the string which precedes any mark $\supset$ is identical with some other string in the set, preceded by (, then the portion of that string which follows the mark $\supset$ referred to may be separately repeated, with the omission of the final mark), and added to the set [Lewis, 1918, p. 357].

Lewis's strictly formal statement of the rule of detachment makes a striking contrast with the informal and inexact version of Whitehead and Russell, which invokes the extraneous notion of truth in setting down the rule.

Post refers several times to the influence of Lewis's formalistic view of logistic systems on his early work [Post, 1921a, p. 165], [Post, 1994, pp. 23,377], and his research at Princeton during his postdoctoral year can be seen as an attempt to work out Lewis's formalistic ideas in the context of the full system of Whitehead and Russell, and to use them in obtaining general results about formal systems.

Post also acknowledged the influence of his thesis advisor, Cassius J. Keyser (1862–1947), on his logical work. This may seem somewhat surprising, in view of the fact that Keyser, although he took part in a debate with Bertrand Russell on the axiom of infinity [Keyser, 1904][Russell, 1904], was best known for his numerous books of popular essays on mathematics and mathematical philosophy. Nevertheless, Post thanks Keyser warmly in the introduction to [Post, 1921a], and his monograph [Post, 1941] is dedicated to Keyser, "in one of whose pedagogical devices the author belatedly recognizes the true source of his truth-table method."

3 THE DOCTORAL DISSERTATION

Post's first published paper in logic [Post, 1921a] is a shortened version of his doctoral dissertation; according to a letter from Post to Hermann Weyl of 1941, the paper was accepted only on condition that its original length be cut by one-third [Davis, 1994, p. xviii].

This article is outstanding for several reasons. Its main mathematical result is the first published proof of completeness and decidability of the propositional fragment of *Principia Mathematica*. (An earlier proof was given by Paul Bernays in his Habilitationsschrift of 1918 at Göttingen, published in abbreviated form in [Bernays, 1926].) The method of the paper is a conscious departure from the methods of Whitehead and Russell, and Post gives a clear statement of the modern metalogical attitude to formal systems. Furthermore, a great deal of Post's most significant later work in logic is present in embryo in this paper.

In the introduction to his paper, Post points out the restricted viewpoint adopted by the authors of *Principia Mathematica*. The aim of Whitehead and Russell is to

develop the logical foundations of mathematics in a fixed axiomatic system (a form of the theory of types). The ideal of the work is to develop mathematics within a logical formalism that is viewed as a universal language. Hence, Whitehead and Russell eschew as far as possible any considerations about their system that go beyond this restricted purpose. Post points out that in this way, they "gave up the generality of outlook which characterized symbolic logic," and goes on to say:

> It is with the recovery of this generality that the first portion of our paper deals. We here wish to emphasize that the theorems of this paper are *about* the logic of propositions but are *not included* therein. More particularly, whereas the propositions of 'Principia' are *particular* assertions introduced for their interest and usefulness in later portions of the work, those of the present paper are about the set of *all* such assertions [Post, 1921a, pp. 163-4].

Post concludes his introduction with a clear explicit statement of his metalogical viewpoint, saying: "We have consistently regarded the system of 'Principia' and the generalizations thereof as purely *formal developments*, and so have used whatever instruments of logic or mathematics we found useful for a study of these developments," referring in a footnote to the chapter of Lewis's book discussed above. Post had thus arrived at a logical method similar to that of the Hilbert school, though in his case, there is no explicit restriction to finitary methods (however, the methods he used were in fact mostly constructive).

The main part of the paper begins with a precise formulation of the propositional calculus of *Principia Mathematica*, inspired by the formulation of Lewis described above. Post then immediately introduces a truth-table development. Denoting the truth-value of a proposition by + if it is true and by − if it is false, he gives the now familiar truth-tables for the primitive connectives

p	$\sim p$
+	−
−	+

$p,\ q$	$p \vee q$
+ +	+
+ −	+
− +	+
− −	−

and goes on to provide a clear explanation of the truth-table method. He uses the adjective "positive" for tautologies, "negative" for contradictions, and "mixed" for contingent formulas (the terminology of "tautology" and "contradiction" is derived from the independent development of the truth-table method in [Wittgenstein, 1921],[Wittgenstein, 1922]). In a footnote, he remarks that the truth-tables for the primitive connectives are described by Whitehead and Russell [Whitehead and Russell, 1910-13, Vol. 1, pp. 8,115], but not the general notion of truth-table, and points out that the idea is implicit in the expansion technique used by the algebraic logicians Boole, Jevons and Venn [Lewis, 1918, p. 74, p. 175f]. As we have seen above, he later cited his thesis advisor Cassius J. Keyser as a more immediate source of the method.

Post now states his "Fundamental Theorem," the result that a propositional function is assertible in the system if and only if it is positive. This provides both a completeness theorem as well as a decision procedure for the system (Post emphasizes the decision procedure aspect rather than completeness). The proof proceeds by showing that any formula in the system is provably equivalent to one in full disjunctive normal form, that is to say, any function $f(p_1, \ldots, p_n)$ can be proved equivalent to a disjunction of conjunctions of the form $Q_1 \wedge \cdots \wedge Q_n$, where each Q_i is either p_i or $\sim p_i$. Post observes that if $f(p_1, \ldots, p_n)$ is a positive function, then the equivalent full disjunctive normal form is easily seen to be provable (since it must contain all possible conjunctions of the form described above), and hence f itself is provable.

Post remarks that while the conversion to full disjunctive normal form and the corresponding decision procedure were both implicit in the earlier literature of algebraic logic, the significance of his theorem is quite different. Although Schröder had given a version of the truth-table decision procedure in his treatise on the algebra of logic, formal and informal logic are bound together in his development in such a way as to prevent the system as a whole from being completely determined. The paper continues by considering three different generalizations of his results, two of which were the basis of a good deal of his later research.

The first generalization consists of generalizing the truth-table method to an arbitrary finite set of truth-functions as primitive connectives. The basis of $\sim$ and $\vee$ chosen by Whitehead and Russell is functionally complete, since all truth-functions can be expressed in this primitive vocabulary (a fact proved explicitly by Post). However, not all choices of primitive connectives result in a functionally complete system; the basis of $\sim$ and $\equiv$ is an example. Post announces but does not prove his complete classification of all possible classes of truth-functions generated by a finite set of connectives. We discuss this result below in §5.

Post next considers "Generalization by Postulation," a much farther-reaching idea. He envisages sets of postulates for generalized propositional logics, with a finite set of primitive connectives, and a finite set of axioms and rules, including the rule of unrestricted substitution, and goes on to consider generalized versions of completeness and consistency for such systems. The usual definition of consistency does not work here, since such systems may not have a recognizable negation connective; similar remarks apply to the notion of completeness. Post replaces these with the following broader notions. He defines a postulate system to be *consistent* if not all formulas are assertible, and *closed* if the addition of any unprovable formula as an axiom results in an inconsistent system. These notions are now usually referred to as "Post-consistency" and "Post-completeness." Earlier in the paper, Post had shown the propositional calculus of *Principia Mathematica* to be consistent and complete in these senses, as corollaries to his Fundamental Theorem. Using his abstract definitions of consistency and completeness, he proves a few general results about closed (Post-complete) systems, remarking of one of them that it "begins to approximate to the truth-table method." This remark looks forward to the work of his postdoctoral year, in which he tried to obtain general

decidability results using the postulational generalization.

The last section of the paper is given over to the final generalization, many-valued logic. Post replaces his earlier truth-values + and − by m distinct "truth-values" $t_1, t_2, \ldots, t_m$, where m is any positive integer. He generalizes classical negation by replacing with a cyclic permutation of the m truth-values; the generalization of classical disjunction has the higher of its two inputs (assuming the ordering $t_1 \geq t_2 \geq \cdots \geq t_m$). Thus the truth-tables for the generalized connectives are as follows:

p	$\sim_m p$
t_1	t_2
t_2	t_3
$\ldots$	$\ldots$
t_m	t_1

$p,\ q$	$p \vee_m q$
$t_1\ t_1$	t_1
$\ldots$	$\ldots$
$t_{i_1}\ t_{j_1}$	t_{i_1}
$\ldots$	$\ldots$
$t_{i_2}\ t_{j_2}$	t_{j_2}
$t_m\ t_m$	t_m

where $i_1 \leq j_1$ and $i_2 \geq j_2$.

Generalizing his earlier results for two-valued logic, Post shows that his m-valued connectives are functionally complete. He then goes on to consider the m-valued logical systems that result by defining the asserted functions to be those that always take values $t_1, t_2, \ldots, t_\mu$, where $0 < \mu < m$. He provides a brief sketch of the method used to give a complete set of postulates for the resulting systems. Jan Łukasiewicz had proposed the idea of a three-valued logic slightly earlier than Post [Łukasiewicz, 1920]. Łukasiewicz was inspired by the idea of a non-Aristotelian logic, in which the third truth-value would represent the notion of indeterminacy, associated with propositions about future events. Post seems mostly interested in his generalization as pure mathematics, remarking that "whereas the highest dimensioned intuitional point space is three, the highest dimensioned intuitional propositional space is two." He goes on to say, though, that just as higher dimensional spaces in geometry can be interpreted by using some other element than point, so the higher dimensioned logical spaces can be interpreted by using elements other than propositions. This idea is carried out in the last section of the paper by interpreting the propositions of m-valued logic as vectors $(p_1, p_2, \ldots, p_{m-1})$ of two-valued propositions, with the proviso that if one proposition is true, then all those that follow are true.

Post himself did not continue his work in many-valued logic, but later mathematicians such as D.L. Webb, Alan Rose and A.L. Foster continued the development of these systems. The algebraic version of Post's systems, the theory of "Post algebras," is investigated in papers such as [Rosenbloom, 1942], [Epstein, 1960], [Traczyk, 1963], [Traczyk, 1964].

4 THE POSTDOCTORAL YEAR

During the academic year 1920-21, Post held a Procter Fellowship at Princeton University (Alan Turing also held a Procter Fellowship 1937-38). It was during this year that he made his most dramatic discoveries in logic, anticipating by a decade or more the work of Gödel, Church, Turing and others on incompleteness and undecidability. Unfortunately, for reasons we discuss below, Post failed to publish his ideas until after other authors had caught up with his groundbreaking work.

He started from the postulational generalization of [Post, 1921a], hoping to obtain general results about the decidability of postulate systems that are abstract versions of the formalism of classical propositional logic. The simplicity of such systems encouraged him in the belief that such results were obtainable, and in addition, he had several positive results to his credit. He had solved the decision problem in the important special case of two-valued propositional logic, and in addition had provided a complete classification of all possible subsystems of this logic based on two-valued connectives. Finally, he had proved the decidability of all abstract propositional logics where all of the connectives were unary (one-place operators), a result published in the abstract [Post, 1921b].

Post's hope was that progress on the decision problem for such generalized propositional calculi (which he called the "general finiteness problem") was possible because of their simple and transparent mathematical structure. His ideas appear nowadays wildly optimistic, but we have the benefit of hindsight; Jacques Herbrand, in a paper surveying the work of the Hilbert school [Herbrand, 1930], expressed a similarly optimistic view towards the decision problem for first-order logic (known as the *Entscheidungsproblem*).

Post published nothing in logic between [Post, 1921b] and [Post, 1936]. Nevertheless, we know of his work of 1920-21 because of the retrospective survey he wrote in 1941 [Post, 1965], on which this section is based. This article falls into two parts. In the first part, Post demonstrates a series of reductions through which the decision problem, or "general finiteness problem" is reduced to a system of an extremely simple type, which he calls a system in "normal form." In the second part, the application of a diagonal argument leads him to the conclusion that the decision problem for systems in normal form is unsolvable, together with the incompleteness results following from this conclusion.

Post defines the notion of reduction as follows: "A system S_1 is reduced to a system S_2 if a 1-1 correspondence is (effectively) set up between the enunciations of S_1, and certain of the enunciations of S_2, so that an enunciation of S_1 is asserted when and only when its correspondent in S_2 is asserted" [Post, 1965, p. 351], [Post, 1994, p. 382]. (This definition is in fact Post's later interpretation of his tacit assumptions; at the time, he did not have a formal definition.) Post uses the word "enunciation" to denote a formula of a system. The most important property of reduction is that if we can solve the decision problem for a system S_2, and S_1 can be reduced to S_2, then the decision problem for S_1 is also solvable.

Post begins from the formulation of a generalized propositional logic described in the previous section, calling it "Canonical Form A." He shows that any system in canonical form A can be reduced to a second form, "Canonical Form B." In Canonical Form B, the unrestricted rule of substitution present in Form A is replaced by a much weaker version. Post shows, however, that the effect of the substitution rule can be obtained by using the idea of axiom schemes. For example, in formulating the propositional calculus, we can take as axioms all formulas of the form $(\sim A \vee (\sim B \vee A))$. This method of dispensing with the rule of substitution is usually attributed to John von Neumann, who employed it in [von Neumann, 1927]. In the conventional formulation, this involves an infinite set of axioms. Post, on the other hand, requires his rules to be finite, so he adds an axiomatization of the notion of formula, or "enunciation," in his terminology. To achieve this, he adds a primitive function $e(\cdot)$ to his system, whose intended interpretation is "P is an enunciation," together with rules such as: $e(P), e(Q) \Longrightarrow e((P \vee Q))$.

Post's next reduction shows that the entire first-order fragment of *Principia Mathematica* can be reduced to canonical form B. Considerable complications are induced by the fact that different types of variable are involved here, while canonical form B only allows one type of variable. A corollary of this work is that if the decision problem for all systems in canonical form B were solvable, then the Hilbertian *Entscheidungsproblem* would be solvable. More important from Post's point of view was the fact that it appeared that all of *Principia Mathematica* could in a similar fashion be reduced to a system in canonical form B. Hence, in spite of its apparently modest appearance, canonical form B already encapsulates the full complexity of the whole system of Whitehead and Russell, at least as far as the decision problem is concerned.

At this point in the narrative of his postdoctoral research, Post introduces a digression into the combinatorial problem of 'tag,' a seemingly innocuous problem that arose early in his research. Let us suppose that we are given a finite set of primitive symbols, say the set of numerals $0, 1, \ldots, \mu$ for simplicity. Then a *tag system* is given by a finite set of rules of the form $0 \to \sigma_0, \ldots, \mu \to \sigma_\mu$, where $\sigma_0, \ldots, \sigma_\mu$ is a sequence of finite strings of primitive symbols, possibly empty, together with a positive integer ν. These rules are applied as follows. Given a non-empty sequence B of symbols, append to the end of B the string associated with the first symbol in B, then delete the first ν symbols (all, if there are less than ν symbols). For example, consider the tag system with the primitive symbols $0, 1$, the rules $0 \to 00, 1 \to 1101$ and $\nu = 3$ (a system that Post himself found intractable). Then if we start with the sequence 10, the rules successively produce $101, 1101, 11101, 011101, \ldots$. The problem of tag is to decide, given an initial string, whether or not the process terminates in the empty string or not. One can think of this as a process in which one end of the string, advancing at a constant rate, is trying to catch up with the other end of the string – hence the colourful name of 'tag,' suggested by Post's fellow-student and colleague B.P. Gill.

Post came upon the problem in the following way. Early in his work, he had discovered a method that, given two formulas in a system in canonical form A

(that is to say, two terms of first-order logic), determines whether or not there is a set of substitutions that makes the two expressions identical. He called this method the "the L.C.M. process" – this is Post's name for the unification algorithm that plays a very important part in automated theorem proving in first-order logic, and was described in print in the groundbreaking paper on the resolution method [Robinson, 1965]. (Post's "least common multiple" appears to be an alternative name for Robinson's "most general unifier.") In attempting to extend the algorithm to higher-order logic, Post encountered the problem of tag. An effort to extend the result of [Post, 1921b] to more general systems led again to the same problem. Thus this combinatorial problem in symbol manipulation seemed like an essential stepping stone towards his hoped-for general results on decidability, and Post made it the major project of his postdoctoral year.

It should come as no surprise that the two problems mentioned above are unsolvable. The problem of tag was proved unsolvable in [Minsky, 1961] – an attractive exposition of this result is in the book [Minsky, 1967]. As for the higher-order unification problem, the second-order case was proved unsolvable in [Goldfarb, 1981]. At the initial stage of his research, though, Post did not have an inkling of unsolvable problems, but instead hoped that the simple and primitive character of tag systems would lead to equally simple decision procedures. This hope led rapidly to disillusionment, as he found that problems of ordinary number theory cropped up in his attempts at solving the problem of tag, disappointing his hope that "the known difficulties of number theory would, as it were, be dissolved in the particularities of this more primitive form of mathematics" [Post, 1965, p. 373], [Post, 1994, p. 398]. Although the solution of the general problem of "tag" appeared hopeless (and hence the entire program of the solution of decision problems), Post had assumed that it was a very minor, if essential, stepping stone towards his broader program. However, in the late summer of 1921, Post found a further series of reductions that led to a final canonical form very close to the apparently special form of "tag." We now describe this final series of reductions.

Canonical form B assumes that there are infinitely many variables, and that the formation rules for formulas are given in terms of functional operations $f(p_1, \ldots, p_n)$. Post begins his series of reductions by reducing systems in form B to systems in canonical form C, in which all formulas are built from a finite alphabet Σ, and the formulas consist simply of the set Σ^* of all finite strings built from symbols in Σ. A system in canonical form C has a finite set of strings as primitive assertions, and a finite set of rules, or "productions" having the form $\sigma_1, \ldots, \sigma_k \Longrightarrow \tau$, where $\sigma_1, \ldots, \sigma_k, \tau$ are strings composed of the symbols in Σ, together with added symbols from a set of auxiliary symbols $\Delta = \{P_1, P_2, \ldots, P_m, \ldots\}$ disjoint from Σ, and every auxiliary symbol in τ appears in one of $\sigma_1, \ldots, \sigma_k$. In an application of the rule, the auxiliary symbols in Δ are considered as syntactic variables ranging over arbitrary strings (possibly empty).

As an example of such a system, consider the alphabet $\{a, b\}$, and let the primitive assertions consist of the strings a, b, aa, bb. The two rules of the system are:

$$P_1 \Longrightarrow aP_1a, \;\; P_1 \Longrightarrow bP_1b.$$

It is easy to see that the assertions of this system consist of all palindromes in the alphabet $\{a, b\}$, that is to say, all non-empty strings that read the same forward as backwards.

As a second example, consider the alphabet $\{p, \sim, \vee, (,)\}$, with the single primitive assertion p, and the rules:

$$P_1 \Longrightarrow \sim P_1, \;\; P_1, P_2 \Longrightarrow (P_1 \vee P_2).$$

Then the assertions of this system are exactly the formulas of the propositional calculus of *Principia Mathematica* containing the single variable p.

Post's next move is to give a series of four successive reductions, in the course of which the decision problem is reduced to that of systems of the very simple Post normal form. A system of canonical form C is said to be in *normal form* if it has a single primitive assertion, and all productions are of the form $\sigma P \Longrightarrow P\tau$. Thus the general decision problem, even for a system as apparently complicated as *Principia Mathematica*, can be reduced to that for a system of rules that can only operate on strings by first deleting an initial substring, then tacking on a new string at the end. This amazing result was described by Marvin Minsky as "one of the most beautiful theorems in mathematics" [Minsky, 1967, p. 240]. Post's proof of the theorem is not inherently difficult, but it is rather lengthy and not very easy to follow; a clear exposition of the theorem, following the lines of Post's original construction, with numerous helpful examples, is provided in Chapter 13 of [Minsky, 1967]. An alternative approach to the proof, based on the theory of Turing machines and semi-Thue systems, is given in Chapter 6 of [Davis, 1958].

The normal form is very close to that of tag systems, since in both cases, the rules involve removing a string of symbols at the beginning, and tacking on a string at the end. The simplicity of the final normal form, together with his earlier success in solving the decision problem for a simple case of canonical form A, aroused Post's hopes anew. However, as he said later, "just when hope was thus renewed for a solution of the general finiteness problem, a fuller realization of the significance of the previous reductions led to a reversal of our entire program" [Post, 1965, p. 402][Post, 1994, p. 418].

4.1 The Anticipation

The reduction of the whole of first-order logic to canonical form B, and the subsequent reduction of systems in canonical form B to systems in normal form made it seem very likely that the whole of *Principia Mathematica* could be reduced to a system in normal form (although Post never in fact worked out all the details of this reduction).

In spite of its apparently great deductive power, a system as complicated as *Principia Mathematica* can be presented as a set of elementary rules for rewriting strings of symbols. This led Post to a far-reaching generalization [Post, 1994, p. 405], [Post, 1994, p. 420]:

> In view of the generality of the system of *Principia Mathematica*, and its seeming inability to lead to any other generated sets of sequences on a given set of letters than those given by our normal systems, we are led to the following generalization.
>
> *Every generated set of sequences on a given set of letters* $a_1, a_2, \ldots, a_\mu$ *is a subset of the set of assertions of a system in normal form with primitive letters* $a_1, a_2, \ldots, a_\mu, a'_1, a'_2, \ldots, a'_\mu$, *i.e., the subset consisting of those assertions of the normal system involving only the letters* $a_1, a_2, \ldots, a_\mu$.

The italicized statement in the quotation equates an intuitive notion, the notion of sets of sequences that are generated by finite means, with a precisely defined mathematical notion. It is analogous to Church's thesis [Church, 1936] identifying functions that are computable in the intuitive sense with general recursive (equivalently, Turing-computable) functions. Post's hypothesis, by analogy, has been called "Post's thesis" [Davis, 1982, p. 21]; under some plausible assumptions, it is in fact equivalent to Church's thesis.

Post now observed that a simple application of Cantor's diagonal argument leads immediately to an apparent counter-example to his thesis. Let us represent sets of positive integers as sets of strings on a single letter 'a', where a string of length n represents the integer n. It is easy to see that we can effectively enumerate all possible normal systems generating such sets of strings. But now the diagonal method apparently leads to a generated set of strings that is not one of those in the enumeration. That is to say, let the diagonal set D be defined by the prescription: a string $a \ldots a$ of length n is in D if and only if it is not generated by the nth normal system in the enumeration. This appears to contradict Post's thesis.

However, a contradiction would only arise if D were in fact a generated set. D would be generated by a finite process, if we make the assumption that the decision problem is solvable for the class of normal systems. What we have shown, in fact, is that this hypothesis together with Post's thesis, leads to a contradiction. Post already had a suspicion that there were inherent difficulties in his quest for general decision methods, based on his difficulties with the problem of tag. However, no difficulties appeared to stand in the way of his thesis. Hence, he held fast to his thesis, concluding that the decision problem for normal systems was in fact unsolvable. He recorded the philosophical influences that led him to the resolution of the contradiction as follows [Post, 1965, p. 407],[Post, 1994, p. 421]:

> In thus resolving this dilemma, the writer was greatly influenced by having heard, not long before, of Brouwer's rejecting the law of the excluded middle. This revolution in the writer's thought was largely energized by the immediately prior reading of Poincaré's *Foundations of Science*.

The informal sketch of the unsolvability of the decision problem for normal systems given above depends on the assumption of Post's thesis, which, since it

equates an informal notion with a mathematical notion, cannot be considered a mathematical hypothesis. However, we can restate the result in such a way as to make into a purely mathematical theorem.

Post concludes the mathematical part of his 1941 paper with a sketch of this restatement. He begins by observing that there is a universal normal system in the following sense. There is a normal system K that has the following property: For every normal system S and formula P in the vocabulary of S, there is a formula (S, P) of K so that (S, P) is assertible in K if and only if P is assertible in S. Post's system K, which he calls *the complete normal system*, is thus a complete inference system for assertions of the form "The string P is assertible in the normal system S," and is the analogue in the realm of normal systems of Turing's universal computing machine [Turing, 1936].

Although the normal systems are abstract versions of formal axiomatic systems, Post observed that we can also use them to represent decision procedures. Suppose that we are given a normal system M, and that in addition, there is a normal system M' that contains all of the primitive symbols of M, and in addition a new letter b (possibly with some other additional symbols). Assume that if P is a formula of M, that P is an assertion of M' if and only if it is an assertion of M, and that bP is an assertion of M' if and only if it is not an assertion of M. Then there is a decision procedure for M, since to decide whether or not a formula P is an assertion of M or not, we can systematically enumerate all of the assertions of M', and observe whether P or bP is an assertion of M'. (This idea is the source of Post's later theorem [Post, 1944, p. 290] that a set of positive integers is recursive if and only if both it and its complement are recursively enumerable.) If such a system M' exists, then Post says that there exists a *finite-normal-test* for the system M [Post, 1965, pp. 412-413], [Post, 1994, p. 425].

In the informal sketch given above of the result that there is no general decision procedure for normal systems, we can replace the appeal to Post's thesis by restricting our attention to normal systems from the outset, and by replacing the informal notion of decision procedure by the concept of finite-normal test. We can thus prove the fundamental result: *There is no finite-normal-test for the complete system K.* A rigorous proof of this result depends on a formal version of the diagonal method sketched above.

Throughout his research of 1920-21, Post concentrated on the problem of decision procedures. However, his idea of representing decision procedures as complete formal theories, as in his notion of "finite-normal-test," leads immediately to incompleteness results for formal systems.

The complete system K is able to give correct positive answers to all questions of the form: "Is the formula P an assertion of normal system S?" Could there be a system that in addition gives correct negative answers to all such questions? The fundamental theorem above indicates that such a system cannot exist. Let L be a normal system that includes all the primitive letters of the complete normal system K among its primitive letters, together with another primitive letter b. Furthermore, let us assume that for any normal system S and formula P of S,

(S, P) is an assertion of L if and only P is an assertion of S, while if $b(S, P)$ is an assertion of L, then P is not an assertion of S. Any such L, Post calls a *normal deductive-system* adjoined to K. In other words, a normal deductive-system adjoined to K gives a correct and complete set of positive answers to questions of the above type, while all of the negative answers it gives to such questions are correct.

The same argument used to prove the fundamental theorem then demonstrates a basic incompleteness theorem: *No normal deductive-system is complete, so that if L is any such system, there is always a normal system S and formula P of S so that P is not an assertion of S, but $b(S, P)$ is not an assertion of L.* This is an abstract form of Gödel's incompleteness theorem, though stated with respect to normal systems rather than arithmetical propositions. Post concludes with the striking words, *A complete symbolic logic is impossible*, and remarks [Post, 1965, pp. 416-417],[Post, 1994, pp. 428-429] :

> This is an iconoclastic result from the formal logician's point of view since it means that logic must not only in some parts of its description (as in the operations), but in its very operation be informal. Better still, we may write
>
> *The Logical Process is Essentially Creative.*

4.2 Post's failure to publish

In view of the truly remarkable nature of the results that Post achieved in the early 1920s during his postdoctoral year, one may well wonder why he did not publish any of this material until fifteen years had passed. There were certainly a number of external reasons for this. The first outbreak of his manic-depressive illness followed immediately on his great discoveries, and the loss of his academic position at Cornell University, together with his being forced to make a living as a high school teacher, were hardly conducive to research and publication. In the summer of 1924, during a year in which he had some association with Cornell University, Post made some progress on his project, but he was never able to bring it to completion.

Another reason given by Post in later years was the general lack of interest in the United States in logical matters. In the covering letter that he sent in 1941 with [Post, 1965] to Hermann Weyl, the editor of the *American Journal of Mathematics*, he points out that, for example, the original paper giving the results of [Post, 1941] was returned to him by the editors of the *Annals of Mathematics* at the height of his postdoctoral work "without any editorial commitment, and with a very mixed report from the referee." He continues:

> It therefore seemed to me to be hopeless to seek publication of Part One of the present paper. And without it, the then revolutionary Part

> Two would have seemed but idle chatter. An attempt to obtain a full proof development was interrupted by ill health and led to a constantly receding date of ultimate publication [Davis, 1994, p. xviii].

This last quotation hints at an internal reason for Post's failure to stake his claim to the incompleteness and undecidability results in time. Kurt Gödel, in the famous incompleteness paper [Gödel, 1931] that led to most of the later research on the limitations of formal systems, set himself a precise but restricted goal. He showed that a powerful, but quite specific, axiomatic system for mathematics, the simple theory of types with the natural numbers as a ground type and the Peano axioms for number theory, was incomplete, and furthermore, that any of a very broad class of ω-consistent extensions of this system shared this incompleteness. At the same time, Gödel left somewhat indefinite the class of systems to which his incompleteness results applied, and it was only later, with the publication of [Church, 1936] and [Turing, 1936] that it became clear that the incompleteness results were completely general, applying to any formal system including a minimum of elementary number theory.

Post, on the other hand, set himself the goal of proving his thesis, that the mathematical concept of normal system exactly characterizes the informal notion of a generated set of strings. In the introduction to a strange but fascinating Appendix to [Post, 1965], he explains his plan:

> While the formal reductions of Part I should make it a relatively simple matter to supply the details of the development outlined in §9 and the beginning of §10, that development owes its significance entirely to the universal character of our characterization of an arbitrary generated set of sequences as given in §7. Establishing this universality is not a matter for mathematical proof, but of psychological analysis of the mental processes involved in combinatory mathematical processes. Because these seemed to be sufficiently simple to be exhaustively described, the writer gave up a direct use of *Principia Mathematica* as a partial verification of the characterization in question, planning rather that the incompleteness of the logic of *Principia Mathematica* would be a corollary of the more general result [Post, 1965, p. 418],[Post, 1994, p. 429].

The remainder of the Appendix consists of quotations from a diary that Post kept under the title "Time Accounts" from the spring of 1916 to the spring of 1922 (this diary is not among the Post papers in Philadelphia, and may have been destroyed or lost).

It is of course possible to provide quite convincing plausibility arguments for Church's thesis, as also for Post's thesis, and [Church, 1936] and [Turing, 1936] both contain such heuristic considerations. Turing's arguments are more convincing than those of Church; the article [Sieg, 1997] contains an illuminating detailed analysis of Church's arguments and a comparison with those of Turing. Nevertheless, it seems hopeless to produce an argument that would establish the result with

complete certainty, and Post's search for such an analysis was a quixotic venture doomed to failure. The result was that his results of 1920-21 were superseded by the work of others, a development that undoubtedly caused Post a great deal of anguish.

Post's unpublished paper of 1941 was an attempt to recoup some of his losses. However, the paper was rejected by Hermann Weyl, who explained his decision by saying:

> I have little doubt that twenty years ago your work, partly because of its then revolutionary character, did not find its due recognition. However, we cannot turn the clock back; in the meantime Gödel, Church and others have done what they have done, and the American Journal is no place for historical accounts ... [Davis, 1994, p. xix]

The paper [Post, 1943] is a greatly shortened version of [Post, 1965], containing only the reduction from canonical form C to normal form, although a concluding lengthy footnote summarizes the content of the rejected paper.

On October 29 1938, Post made the acquaintance of Kurt Gödel at a regional meeting of the American Mathematical Society in New York, and spoke with him about his work on absolutely unsolvable problems. On the same day, he wrote Gödel a touching letter in which he said:

> I am afraid that I took advantage of you on this, I hope but our first meeting. But for fifteen years I had carried around the thought of astounding the mathematical world with my unorthodox ideas, and meeting the man chiefly responsible for the vanishing of that dream rather carried me away [Gödel, 2003, p. 169].

On the following day, Post, still excited by his meeting with Gödel, wrote a second letter in which he summarized his work of the 1920s, following the outline of [Post, 1965] while emphasizing the incompleteness of any formal system with respect to the fixed subject matter of provability in normal systems. Explaining his failure to prove Gödel's results at the time, he said:

> May I finally say that nothing that I had done could have replaced the splendid actuality of your proof. For while corollary your theorem may be of the existence of an absolutely unsolvable problem, the absolute unsolvability of that problem has but a basis in the nature of physical induction at least in my work and I still think in any work. Of course with sufficient labor that induction could have gone far enough to include your particular system theorematically. That that could be done for Principia Mathematica I saw then. My only excuse for not doing so – well there are many and having written so much I might add them. Chiefly I thought I saw a way of so analyzing "all finite processes of the human mind" [something of the sort of thing Turing does in his computable number paper] that I could establish the above

> conclusions in general and not just for Principia Mathematica. Secondly that the absolute unsolvability of my problem would not achieve much recognition for others merely on the basis of an incompleteness proof for Principia Mathematica. And lastly while the above general analysis enticed me it seemed foolish to do all the labor involved in the more special Principia Theorem.

Post concluded his letter by apologizing to Gödel for his excitement when they met:

> It was a real pleasure to meet you and I hope my egotistical outbursts have spent themselves with that first meeting and this letter. Needless to say I have the greatest admiration for your work, and after all it is not ideas but the execution of ideas that constitute a mark of greatness [Gödel, 2003, pp. 171-72].

In a brief reply, Gödel reassured Post that he had noticed "nothing of what you call egotistical outbursts in your letters or in the talk I had with you in New York; on the contrary it was a pleasure to speak with you" [Gödel, 2003, p. 173].

In spite of his disappointment in being overtaken by other logicians, Post continued to work in the area of recursion theory, and made some of his most important and influential contributions in the period from 1936 to 1954. Before we take up this thread again, however, it is necessary to make a digression in order to describe Post's work on Boolean clones. Although his fundamental results in this area were already announced in [Post, 1921a], they were finally published twenty years later after major changes in exposition.

5 CLASSIFYING SETS OF BOOLEAN FUNCTIONS

In this section, we discuss Post's basic results on classes of Boolean functions closed under composition. Although his work in this area belongs with the most fundamental results of modern mathematical logic, it has remained somewhat difficult of access.

When Post rewrote his research for publication, following an unfortunate suggestion of a referee [Post, 1941, p. 4], [Post, 1994, p. 256] he adopted the notation of Jevons for logical formulas (his earlier version was in terms of the more familiar truth-tables). Jevons uses capital letters $A, B, C, \ldots$ as Boolean variables (interpreted as classes), while the corresponding lower case letters $a, b, c, \ldots$ represent the complements of these variables [Jevons, 1864]. The addition symbol + represents Boolean union (inclusive disjunction), while the universal and empty classes are symbolized by 1 and 0. As an example of Jevons's notation, the disjunctive normal form expression usually symbolized as $(A \wedge B) \vee (\sim A \wedge B \wedge \sim C)$ is written in Jevons's notation as $AB + aBc$. Post proceeds in his classification by means of what he calls an "expansion" of a logical function, a pair $\alpha : \beta$ of expressions

in disjunctive normal form, expressed in Jevons's notation, the first being a disjunctive normal form of the function itself, the second a disjunctive normal form of its negation. The expansion is not unique; for example, one expansion of the implication function $A \supset B$ is $a + B : Ab$, while another is $AB + aB + ab : Ab$.

Although Jevons's notation is concise, and not difficult to learn, Post's use of it leads to quite unnecessary obscurities. For example, the familiar class of monotone functions in Post's terminology becomes the class of functions "satisfying the $[A : a]$ condition" [Post, 1941, p. 35], [Post, 1994, p. 287]. This idiosyncratic phrase derives from a characterization of monotone functions in terms of their expansions. A Boolean function f is monotone if and only if it satisfies the $[A : a]$ condition: for any two terms in an expansion of f, there is a letter which is capital in the term that is in the first expression and small in the term that is in the second expression.

The result, unfortunately, is a monograph that modern logicians seem to find impossible to read. Post provides a complete and painstaking proof of his major results; reviewing the work in the *Journal of Symbolic Logic*, H.E. Vaughan [Vaughan, 1941] remarked that the book is "self-contained and very clearly written," and noted only four serious misprints. Nevertheless, the contemporary logician Roman Murawski wrote [Murawski, 1998] of Post's characterization of functionally complete sets of connectives (proved in §26 of his monograph):

> Did Post prove this theorem? In [Post, 1941] one finds no proof satisfying the standards accepted today. The reason for that was Post's baroque notation (it was in fact an unprecise adaptation of the imprecise notation of Jevons from his *Pure Logic*, cf. [Jevons, 1864]), other reason was the fact that Post seemed to be simultaneously pursuing several different topics.

Although these accusations of imprecision and lack of rigour are quite unjustified, Murawski's remarks (which echo similar but somewhat more guarded observations in [Pelletier and Martin, 1990, p. 463]) form an eloquent testimony to the unreadability of Post's ill-fated monograph.

In view of the problematic nature of Post's exposition, in most of this section we shall expound his results in unabashedly modern terminology. The exposition below in fact has some substantive differences from Post's original statement of his results; towards the end of this section, we shall return to Post's original version to explain where these differences lie.

5.1 A tour of Post's lattice

Our object of study in this section is the family of all finitary functions on the two-element set $S = \{0, 1\}$; we can identify the elements with truth-values, with 1 as "true," and 0 as "false." An n-place operation on S, for $n > 0$, maps n-tuples of Boolean values into Boolean values. There are 2^{2^n} such functions for each n; familiar logical operations such as $\wedge$ and $\vee$ are among the sixteen 2-place Boolean

functions. (We might also consider the constants 0 and 1 as zero-place functions, but here we prefer to think of them as 1-place constant functions.)

The most basic n-place functions (apart from the constant functions) are the *projections* π_i^n, defined by the equation: $\pi_i^n(x_1, \ldots, x_n) = x_i$. If we think of functions as given by expressions involving variables $x_1, \ldots, x_i, \ldots, x_n, \ldots$, then the projections correspond to the variables themselves. We can construct new functions from old by the operation of *composition*. If f is a k-place operation, and $g_1, \ldots, g_k$ a list of n-place functions, then we define an n-place function h by composition:

$$h(x_1, \ldots, x_n) = f(g_1(x_1, \ldots, x_n), \ldots, g_k(x_1, \ldots, x_n)).$$

We define a family of Boolean functions to be a *clone* if it contains all the projection functions and is closed under composition (Post uses a more general notion of "iteratively closed class," described below in §5.2).

Figure 2 shows the inclusion diagram of all Boolean clones. Post's results are perhaps most easily understood in terms of a commentary on this diagram. The first and most obvious feature of the diagram is its left-right symmetry. This symmetry arises from the phenomenon of duality. If $f(x_1, \ldots, x_n)$ is a Boolean function, then the function g defined by the equation

$$g(x_1, \ldots, x_n) = \sim f(\sim x_1, \ldots, \sim x_n)$$

is the *dual* of f; the dual of the dual of f is just f itself. Equivalently, the truth-table of the dual of f is obtained from the truth-table of f by replacing 0 by 1 and 1 by 0. For example, $\wedge$ is the dual of $\vee$, and vice versa; the dual of the implication operator $(x \supset y)$ is the subtraction operator $y \smallsetminus x$, or equivalently, $(y \wedge \sim x)$. A function is *self-dual* if it is equal to its dual. If C is a clone, then the class of functions dual to those in C is also a clone, so that to every clone in our diagram there is a corresponding dual clone, explaining the left-right symmetry. The fourteen systems in the centre of the diagram are self-dual classes (that is to say, they are their own duals).

There are only countably many Boolean clones, and Post's classification gives a complete, explicit description of all of them. Figure 2 shows a central section, consisting of 30 "sporadic" clones, together with eight infinite descending chains, four on the left and four on the right. The supplementary Figure 3 shows some of the rather intricate pattern of containments in the central part of the lattice. Every one of the clones is finitely generated, that is to say, there is an explicitly describable finite set G of functions so that every function in the clone can be defined from functions in G. In the remaining part of this section, we shall take a tour of Post's lattice, giving a description of all of the clones and their generators.

At the bottom of the lattice, we find six classes generated by negation and constants: $R_1 = \{\emptyset\}$ (the set of all projections), $R_4 = \{\sim\}$, $R_6 = \{1\}$, $R_8 = \{0\}$, $R_{11} = \{1, 0\}$, $R_{13} = \{\sim, 1\}$. Here, the notation $\{\sim, 1\}$ (for example) denotes the smallest clone containing the functions $\sim$ and 1.

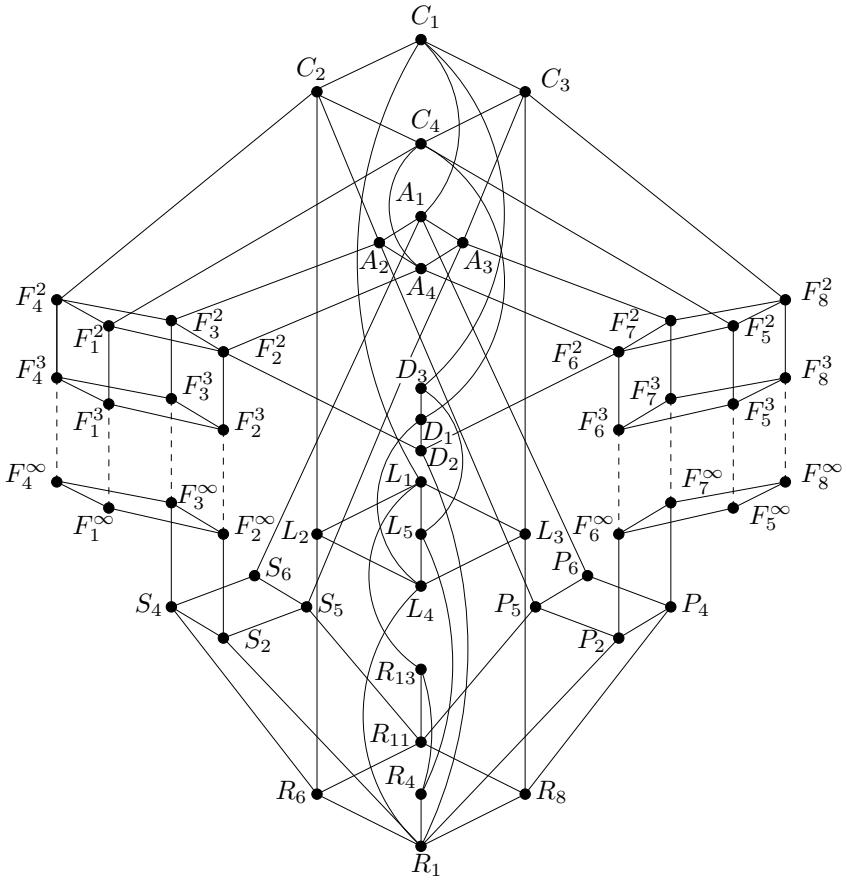


Figure 2. Post's lattice

Proceeding up the lattice, we encounter two groups of four on left and right, and one group of five in the middle. On the left, we have four classes generated by disjunction and the constants: $S_2 = \{\vee\}$, $S_4 = \{\vee, 1\}$, $S_5 = \{\vee, 0\}$ and $S_6 = \{\vee, 0, 1\}$. On the right, the dual classes are: $P_2 = \{\wedge\}$, $P_4 = \{\wedge, 0\}$, $P_5 = \{\wedge, 1\}$, $P_6 = \{\wedge, 0, 1\}$. In the middle, there are five classes describable in terms of the biconditional $\equiv$ and its dual $\oplus$ (that is to say, addition modulo 2). If we write $\oplus_3$ for the three place operation $x \oplus y \oplus z$, then we have: $L_1 = \{\equiv, \sim\}$, $L_2 = \{\equiv\}$, $L_3 = \{\oplus\}$, $L_4 = \{\oplus_3\}$, $L_5 = \{\oplus_3, \sim\}$. The class L_1 can also be described abstractly as the set of all *linear* Boolean functions, that is to say, those expressible in the form $x_1 \oplus \cdots \oplus x_n \oplus c$, where c is a constant (0 or 1).

At this point, it is convenient to make a digression to introduce some definitions. We define two infinite sequences, $c_2, c_3, \ldots, c_n, \ldots$ and $d_2, d_3, \ldots, d_n, \ldots$ of

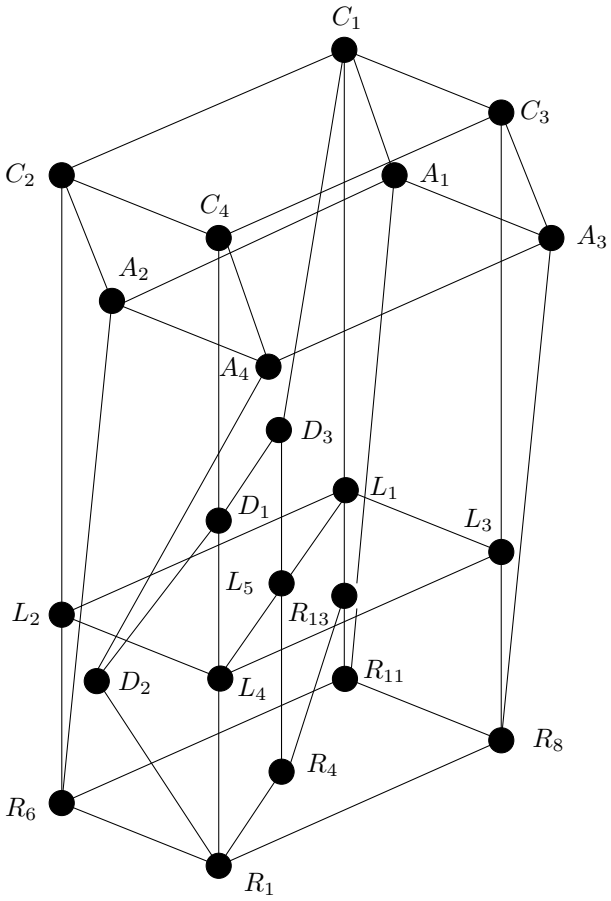


Figure 3. Post's lattice: the central part

functions, where c_n and d_n each have $n+1$ input variables. For $n \geq 2$, we define:

$$c_n(x_1, x_2, \ldots, x_{n+1}) = \bigwedge_{1 \leq i \leq n+1} (x_1 \vee \cdots \vee \widehat{x_i} \vee \cdots \vee x_{n+1}),$$

$$d_n(x_1, x_2, \ldots, x_{n+1}) = \bigvee_{1 \leq i \leq n+1} (x_1 \wedge \cdots \wedge \widehat{x_i} \wedge \cdots \wedge x_{n+1}),$$

where the notation "$\widehat{x_i}$" means that the variable x_i is omitted from the term in question. For example, function d_2 is given by the definition

$$d_2(x_1, x_2, x_3) = (x_2 \wedge x_3) \vee (x_1 \wedge x_3) \vee (x_1 \wedge x_2).$$

The function d_n belongs to the family of *threshold functions*, a family that plays an important role in theoretical computer science [Vollmer, 1999]. The

threshold function T^n_m, for $m \leq n$, is the n-place function defined by the condition: $T^n_m(x_1, \ldots, x_n) = 1$ if and only if $x_1 + \cdots + x_n \geq m$ (that is to say, at least m of the inputs $x_1, \ldots, x_n$ are equal to 1). For example, $(x \vee y) = T^2_1$, and $(x \wedge y) = T^2_2$. With this definition, we have $d_n = T^{n+1}_n$. The function c_n satisfies the dual condition: $c_n(x_1, \ldots, x_n) = 1$ if and only if $x_1 + \cdots + x_n > 1$. It is easy to see that $c_2 = d_2$. The functions c_n and d_n are duals of each other, so that the function c_2 is self-dual.

In the centre of the diagram is a group of three classes, $D_2 = \{d_2\}$, $D_1 = \{d_2, \oplus_3\}$ and $D_3 = \{d_2, \oplus_3, \sim\}$. D_3 is the class of all self-dual functions; hence there are exactly seven classes consisting entirely of self-dual functions.

Immediately above D_3 lies a group of four classes, A_1, A_2, A_3 and A_4. The top element in this group, A_1, is the important class of all monotone functions. If we order the truth-values by setting $0 < 1$, then n-tuples of truth-values can be ordered by defining $\langle x_1, \ldots, x_n\rangle \leq \langle y_1, \ldots, y_n\rangle$ if and only if $x_i \leq y_i$, for all i, $1 \leq i \leq n$. An n-place function f is *monotone* if for all $x_1, \ldots, x_n, y_1, \ldots, y_n$,

$$\langle x_1, \ldots, x_n\rangle \leq \langle y_1, \ldots, y_n\rangle \Rightarrow f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n).$$

The generators of these classes are given by $A_1 = \{\wedge, \vee, 0, 1\}$, $A_2 = \{\wedge, \vee, 1\}$, $A_3 = \{\wedge, \vee, 0\}$ and $A_4 = \{\wedge, \vee\}$.

Proceeding to the top of the lattice, we find C_1, the class of all Boolean functions, C_2, the class of all functions that preserves 1, in the sense that if f is an n-place function in C_2, then $f(1, 1, \ldots, 1) = 1$, C_3, the class of all functions preserving 0, and C_4, the class of all functions preserving both 1 and 0. For generators, we have $C_1 = \{\vee, \sim\}$, $C_2 = \{\wedge, \supset\}$, $C_3 = \{\vee, \diagdown\}$, $C_4 = \{x \wedge (y \vee \sim z), \vee\}$.

Finally, we have to describe the eight infinite descending chains arranged symmetrically on each side of the lattice. Starting with the four chains on the left, we can describe them using the following sets of generators, where $2 \leq m < \infty$:

$$F^m_1 = \{x \vee (y \diagdown z), c_m\}, F^\infty_1 = \{x \vee (y \diagdown z)\};$$

$$F^m_2 = \{x \vee (y \wedge z), c_m\}, F^\infty_2 = \{x \vee (y \wedge z)\};$$

$$F^m_3 = \{c_m, 1\}, F^\infty_3 = \{x \vee (y \wedge z), 1\};$$

$$F^m_4 = \{\supset, c_m\}, F^\infty_4 = \{\supset\}.$$

The four chains on the righthand side of the lattice are the duals of those on the left. Thus, a set of generators for each class is given by the duals of the generators of the corresponding dual class:

$$F^m_5 = \{x \wedge (y \vee \sim z), d_m\}, F^\infty_5 = \{x \wedge (y \vee \sim z)\};$$

$$F^m_6 = \{x \wedge (y \vee z), d_m\}, F^\infty_6 = \{x \wedge (y \vee z)\};$$

$$F^m_7 = \{d_m, 0\}, F^\infty_7 = \{x \wedge (y \vee z), 0\};$$

$$F^m_8 = \{\diagdown, d_m\}, F^\infty_8 = \{\diagdown\}.$$

It is possible to give a more structural description of the eight infinite chains by giving properties that characterize the functions in each clone. We generalize the property of being 0-preserving that picks out the functions in the class C_3. Define a function f to be *doubly 0-preserving* if, given two n-tuples of truth-values, $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ such that $f(x_1, \ldots, x_n) = f(y_1, \ldots, y_n)$, there is a j, where $1 \leq j \leq n$, so that $x_j = y_j = 1$. More generally, define f to be *m-tuply 0-preserving* if whenever $x_1^1, \ldots, x_n^1$, $x_1^2, \ldots, x_n^2$, $\ldots$, $x_1^m, \ldots, x_n^m$ is a sequence of length m, consisting of n-tuples of truth-values such that $f(x_1^1, \ldots, x_n^1) = f(x_1^2, \ldots, x_n^2) = \cdots = f(x_1^m, \ldots, x_n^m) = 1$, there is a j, where $1 \leq j \leq n$, so that $x_j^1 = x_j^2 = \cdots = x_j^m = 1$. A function is *infinitely 0-preserving* if it is m-tuply 0-preserving for all m. This terminology, taken from [Pippenger, 1997, p. 22], is justified by the fact that a function is 0-preserving just in case it is singly 0-preserving.

The class F_8^m consists of all m-tuply 0-preserving functions, F_7^m results by restricting such functions to be monotone. F_5^m consists of all m-tuply 0-preserving functions that are also 1-preserving, while F_6^m is the subclass of all monotone functions in F_5^m. Finally, the classes F_5^∞, F_6^∞, F_7^∞, F_8^∞ are characterized in exactly the same way, except that the crucial property is that of being infinitely 0-preserving.

The classes F_1^m, F_2^m, F_3^m, F_4^m and F_1^∞, F_2^∞, F_3^∞, F_4^∞ are characterized in the same way as their dual counterparts, except that the dual property, of being m-tuply 1-preserving is employed.

The descriptions of the classes in terms of their generators is adapted from the paper [Lyndon, 1951] (see also [Böhler *et al.*, 2003]). In that paper, Lyndon provides a complete axiomatization for the equational logics corresponding to each of Post's classes.

5.2 Post's iterative classes

In the exposition of Post's results, following the contemporary trend, we have included all projection functions in our functionally closed classes. This is a natural approach, since (for example) we can consider the formula $\sim x$ as expressing a function on any number of variables, simply by regarding the variables other than x as irrelevant. In other words, when we include an n-place function in a class, then we also include all those functions that arise from it by adding irrelevant variables.

Post, however, did not make this assumption, and as a result, his classification includes twenty iteratively closed classes that are not clones. We first need to define Post's own notion of iteratively closed class. Let π be a map from $\{1, \ldots, n\}$ onto $\{1, \ldots, k\}$, where $k \leq n$. Define $\{0,1\}^n/\pi$ to be the class of all $x \in \{0,1\}^n$ satisfying the condition that for $i, j \leq n$, if $\pi(i) = \pi(j)$ then $x_i = x_j$. If $x \in \{0,1\}^k$, then there is a unique vector $y = \pi^{-1}(x)$ in $x \in \{0,1\}^n/\pi$ so that for all i, $1 \leq i \leq n$, $y_i = x_{\pi(i)}$. If f is a Boolean function defined on $\{0,1\}^n$, then we define the function f^π on $\{0,1\}^k$ by setting $f^\pi(x) = f(\pi^{-1}(x))$. We now define a family

F of Boolean function to be *iteratively closed* if it satisfies the conditions:

1. If f is an n-place function in F and π a map from $\{1, \ldots, n\}$ onto $\{1, \ldots, k\}$, where $k \leq n$, then f^π is in F;

2. If $f(x_1, \ldots, x_i, \ldots, x_k)$ is a k-place function in F, and $g(x_1, \ldots, x_n)$ an n-place function in F, then the $k + n - 1$-place function h defined by

$$h(x_1, \ldots, x_{k+n-1}) = f(x_1, \ldots, g(x_i, \ldots, x_{i+k-1}), x_{i+k}, \ldots, x_{k+n-1})$$

   is also in F.

That is to say, Post defines a non-empty class to be iteratively closed if it is closed under permutation and identification of variables, and also under composition. We now give a short description of the twenty iteratively closed classes that are not clones.

There are nine classes consisting entirely of 1-place functions. These are $O_1 = \{\pi^1_1\}$, $O_2 = \{1\}$, $O_3 = \{0\}$, $O_4 = \{\pi^1_1, \sim\}$, $O_5 = \{\pi^1_1, 1\}$, $O_6 = \{\pi^1_1, 0\}$, $O_7 = \{1, 0\}$, $O_8 = \{\pi^1_1, 1, 0\}$ and $O_9 = \{\pi^1_1, 1, 0, \sim\}$. Since Post does not consider the empty class to be iteratively closed, his containment diagram has no least element, and hence is not a lattice, since O_1, O_2 and O_3 are incomparable minimal elements.

Seven classes consist of sets of functions reducible to 1-place functions, that is to say, all functions in these classes are obtained by adding irrelevant variables to 1-place functions. Let 1^* be the class of all n-place constant functions taking the value 1, and 0^* the class of all n-place constant functions taking the value 0. Then we have the additional iterative classes $R_2 = \{1^*\}$, $R_3 = \{0^*\}$, $R_5 = \{\pi^1_1, 1^*\}$, $R_7 = \{\pi^1_1, 0^*\}$, $R_9 = \{1^*, 0^*\}$, $R_{10} = \{\pi^1_1, 1^*, 0^*\}$ and $R_{12} = \{\pi^1_1, 1^*, 0^*, \sim\}$.

Finally, there are four classes defined in terms of conjunction and disjunction. $S_1 = \{\pi^1_1, \vee\}$, $S_3 = \{\pi^1_1, 1^*, \vee\}$, $P_1 = \{\pi^1_1, \wedge\}$ and $P_3 = \{\pi^1_1, 0^*, \wedge\}$.

5.3 Applications of Post's classification

Post's classification, together with the accompanying containment diagram, encodes an enormous quantity of information. Many results, some of them far from obvious, can be read off immediately from the diagram.

As an example of such a result, consider the following assertion: If S is an infinite set of Boolean functions, then there is a finite set S_0 contained in S so that every function in S can be defined from functions in S_0, together with the projection functions. This does not seem at all obvious. However, it follows immediately by inspecting Post's lattice. The lattice, although it contains infinite descending chains, contains no infinite ascending chains, and the result follows easily from this fact.

Post's lattice also makes it easy to see which class is generated by a given finite set of Boolean functions. In particular, we can readily give a criterion for a set of functions F to be functionally complete, so that all functions are definable from F. To see how this works, let us examine the top element C_1 of Post's lattice. The

class C_1 covers exactly five other classes; that is to say, exactly five classes, namely C_2, C_3, L_1, A_1, D_3, lie immediately below C_1 in the diagram. It follows from this that if F is a set of Boolean functions, then it is functionally complete if and only if for every one of these five classes, there is a function in F that is not in the class.

The last sentence of the previous paragraph can be rephrased to give Post's criterion for functional completeness. A set of functions is functionally complete if and only if it contains (a) a function that does not preserve 0, (b) a function that does not preserve 1, (c) a non-monotone function, (d) a non-linear function, (e) a function that is not self-dual. Since each one of these conditions is easy to check, given the truth-table for a function, the criterion leads to a simple algorithm to determine whether or not a given finite set of functions is functionally complete. A self-contained proof of Post's theorem on functional completeness can be found in [Pelletier and Martin, 1990]. Similar criteria can be given for all the other classes in Post's diagram.

For several decades after the publication of [Post, 1941], Post's classification was consigned to a backwater of logic, and was almost forgotten. However, with the increasing interest in Boolean functions and propositional logic, this situation has changed, and his fundamental results are now frequently employed by theoretical computer scientists. A survey of some of these applications of Post's classification can be found in the excellent expository article [Böhler *et al.*, 2003].

6 RECURSION THEORY AND DECISION PROBLEMS

Although he had been overtaken by the work of other logicians, Post was well placed to contribute to the newly developing fields stemming from the discovery of the incompleteness of formal systems by Gödel, and of an exact mathematical definition of computable functions by Church and Turing. His point of view was substantially different from the approach followed by other logicians, so that his papers of the 1930s and 1940s are highly original, breaking new ground in several directions.

Most of the work in recursion theory in the mid-1930s emphasized computable functions on the natural numbers. Post's earlier work differed in its overall thrust, since he was mainly concerned with formal axiomatic systems, expressed as rules for manipulating strings of symbols. Consequently, Post was able to develop the theory in new directions, emphasizing both generated sets of numbers and problems involving rules for symbolic manipulation.

6.1 Post machines

Post's first publication in logic after a fifteen year gap was [Post, 1936], a short paper in which he describes a computational model very similar to that of the Turing machine. Turing's article was received for publication on May 28 1936, while Post's was received October 7 1936; a footnote to Post's article by the editor, Alonzo Church, informs the reader that "the present article, however, although

bearing a later date, was written entirely independently of Turing's." Post, on the other hand, unlike Turing, had the advantage of having read [Church, 1936], while Turing's logical influences seem to have been confined to the incompleteness results [Gödel, 1931], the Cambridge lectures of Max Newman and the textbook [Hilbert and Ackermann, 1928].

Unlike Turing, Post does not use the word "machine," but instead speaks of a "problem solver or worker" (however, the term "Post machine" [Uspensky, 1983] is now standard terminology for the model). The analogue of a Turing machine's "tape" is provided by a "symbol space" consisting of a two way infinite sequence of spaces or boxes, each of which is either empty or unmarked. The worker is assumed to be performing the following primitive acts:

(a) Marking the box he is in (assumed empty),

(b) Erasing the mark in the box he is in (assumed marked),

(c) Moving to the box on his right,

(d) Moving to the box on his left,

(e) Determining whether the box he is in, is or is not marked.

The *set of directions* for the worker consists of a sequence of numbered instructions, each instruction having one of the following forms:

(A) Perform operation (a), (b), (c) or (d) and then follow direction j,

(B) Perform operation (e) and according as the answer is yes or no correspondingly follow direction number i or direction number j,

(C) Stop.

One box is to be singled out and called the starting point. A problem is given in symbolic form by a finite number of boxes being marked with a stroke, and the output is given in the same form. With these conventions, a set of instructions serves to define a computable function, where the positive integer n is symbolized by a sequence of n consecutive marked boxes.

Post conjectures (correctly) but does not prove that his formulation is equivalent to Church's definition of computable functions. His formulation is strikingly similar to Turing's earlier definition, though Turing's notion of machine configurations or "m-configurations" does not appear in Post's development. This role is played by the numbers assigned to the instructions in Post's sets of directions. Although Turing's notion of computing machine, rather than Post's, has become the standard model in theoretical computer science, Post's formulation is a little more like standard machine code (since programs typically do consist of numbered instructions). To put the matter in computer science terms: Post shows that a computer that can only perform the primitive acts of reading, writing, moving and

following jump instructions conditional on a zero test is a universal computational model.

Post evidently had plans to continue in a series of articles on the topic (as appears in the title of his paper) showing how "Formulation 1" could be extended to broader computational models. He must have been discouraged from this by the appearance of the more elaborate development in [Turing, 1936], and one can sense his disappointment in a remark recorded in a notebook in February 1938:

> *Turing:* In large measure removed what was left in my point of view. His *Finite number of mental states* hypothesis I did not have, and if verified should be a cornerstone in the finite process development [Grattan-Guinness, 1990, p. 82].

Post's notebooks in Philadelphia (see §8) contain some of his work on "Formulation 2." In it, he considered rules operating in a two-dimensional symbol space, somewhat reminiscent of later work on cellular automata, such as Conway's Game of Life [Berlekamp *et al.*, 1982, Chapter 25] [Wolfram, 2002].

Post ends his paper with some remarks on the significance of Church's thesis. He says of his conjecture that his formulation 1 is equivalent to the Gödel-Church development: "We offer this conclusion at the present moment as a *working hypothesis*. And to our mind such is Church's identification of effective calculability with recursiveness." To this last sentence he adds a footnote, in which he says: "Actually the work done by Church and others carries this identification considerably beyond the working hypothesis stage. But to mask this identification under a definition hides the fact that a fundamental discovery in the limitations of the mathematicizing power of Homo Sapiens has been made and blinds us to the need of its continual verification" [Post, 1936, p. 105].

6.2 Recursively Enumerable Sets

In February 1944, Post gave an invited address to the New York meeting of the American Mathematical Society, and the published version of his talk [Post, 1944] is his most influential paper. In it, he sets forth a program to determine the relative complexity of undecidable problems, based on abstract notions of reducibility. As part of this program, he poses a problem, later known as "Post's problem" that was one of the driving forces behind the rapid development of recursion theory in the decades following Post's address.

Since he was addressing a general mathematical audience, Post adopted an informal mode of presentation that contrasts with the rather formal exposition that was standard in the field of recursive functions at that time. He remarks:

> We must emphasize that, with a few exceptions explicitly so noted, we have obtained formal proofs of all the consequently mathematical theorems here developed informally. Yet the real mathematics involved must lie in the informal development. For in every instance the informal "proof" was first obtained; and once gotten, transforming it into

> the formal proof turned out to be a routine chore [Post, 1944, p. 284] [Davis, 1965, p. 305][Post, 1994, p. 462].

Although Post seems to have felt that publication of final results required a more formal mode, in fact his informal expository style has now become the norm in presentations of research in recursion theory.

Post begins his article with a beautifully written survey of the area, stating Church's thesis as the identification of the intuitive concept of effectively calculable function with the technical concept of recursive function. He goes on to explain the intuitive concept of a *generated set* of positive integers, and asserts his own thesis as the identity of this concept with that of recursively enumerable set. The technical definition of recursively enumerable set is given in terms of his own concept of normal system, a set of positive integers being recursively enumerable if it is generated by a normal system with the alphabet $\{1, b\}$, the positive integer n being represented by a string of n 1's; if a set of integers is generated by a set of rules B, then we say that B is a *basis* for the set generated by these rules. He goes on to point out the generality of this concept, since the assertions of an arbitrary formal system of formal logic can (by numerical encoding) be regarded as a recursively enumerable set.

With this identification, the decision problem for a formal system can be seen as the decision problem for a recursively enumerable set. A corollary of Church's thesis is that the intuitive concept of a set with a solvable decision problem is coextensive with the notion of recursive set. Hence, questions of algorithmic decidability can be placed in the abstract framework of recursively enumerable sets of positive integers and their decision problems. The basic result on the existence of unsolvable problems (for which Post references [Church, 1936], [Rosser, 1936] and [Kleene, 1936]) is stated as: *There exists a recursively enumerable set of positive integers which is not recursive.*

Post continues with a version of Gödel's incompleteness theorem, a reworked version of his earlier incompleteness results from [Post, 1965]. The set of all bases $B_1, B_2, \ldots$ can be effectively enumerated, and hence so can the set T of all pairs (B, n), where n is in the set generated by the basis B. This set T is thus the same set as is generated by the complete normal system described in §4.1. If we define $\overline{T}$ as the set of all pairs (B, n) not in T, then $\overline{T}$ is not recursively enumerable, by a diagonal argument. Thus there is no formal logical system that proves a proposition of the form "The integer n is not in the set generated by the basis B" if and only if it is true. Post concludes with words echoing his earlier remarks from his unpublished manuscript of 1941 [Post, 1944, p. 295] [Davis, 1965, p. 316]:

> The conclusion is unescapable that even for such a fixed, well defined body of mathematical propositions, *mathematical thinking is, and must remain, essentially creative.* To the writer's mind, this conclusion must inevitably result in at least a partial reversal of the entire axiomatic trend of the late nineteenth and early twentieth centuries, with a return to meaning and truth as being of the essence of mathematics.

With these words, Post sums up his earlier work on undecidability and incompleteness. The remainder of his paper sets out a program of research that was to determine a great deal of the direction of the new field of recursion theory. The emphasis is now on the classification of recursively enumerable sets, with the measure of complexity provided by notions of reducibility.

In his postdoctoral work at Princeton, Post had used an informal notion of reducibility in relating systems in his various canonical forms. If we translate Post's earlier notion of reducibility, cited above in §4, into the language of set theory, then it amounts to that of *one-one reducibility*. If S and T are sets of positive integers, then S is *many-one reducible* to T if there is a computable function f defined on the positive integers so that n is in S if and only if $f(n)$ is in T. S is *one-one reducible* to T if the function f is injective (one-one).

The intuitive idea of reducibility is that it measures the degree of difficulty of a problem. If a problem P is reducible to a problem Q, then a solution to Q would also lead to a solution of P. Different notions of reducibility arise, depending on the complexity of the reduction process. The reducibility relation is always assumed to be transitive and reflexive, so that we can classify sets in a partial ordering of degrees of unsolvability, where two sets are equivalent under the ordering if and only they are reducible to each other (in this case, they are said to be of the same *degree*). The most restrictive notion of reducibility defined by Post is that of one-one reducibility; later he introduces more and more general notions of reducibility, the most general notion being that of *Turing reducibility*.

Because the complete set K encodes positive answers to all questions of the form "Is the integer n in the set generated by the basis B?", it is not hard to see that every recursively enumerable set is one-one reducible to K. Conversely, it is easy to prove that if a set is many-one reducible to K, then it is recursively enumerable. Hence, the complete set K is of the highest degree of unsolvability relative to one-one reducibility (*1-complete*), and consequently also relative to many-one reducibility (*m-complete*).

Are there degrees of unsolvability of recursively enumerable sets strictly between the degree of recursive sets and that of K? This is the general problem addressed by Post in the later parts of his paper; he answers the question for some of the stronger (that is, more restrictive) concepts of reducibility, but not for the most general notion, Turing reducibility. As an initial step in his quest for intermediate degrees, Post begins by singling out a key property of the complete set K.

Fix a numerical encoding of all of the pairs (B, n). Then the image K of the set T under this encoding is a recursively enumerable set that is not recursive. The diagonal method that proves K not to be recursive is constructive. In Post's terminology, the set K is a *creative* set, that is to say, a recursively enumerable set C for which there exists a recursive function f giving a unique positive integer $n = f(i)$ for each basis B_i for a recursively enumerable set W_i of positive integers such that whenever W_i is a subset of $\overline{C}$ (the complement of C), then n is also in $\overline{C}$, but not in W_i. This key property of the complete set K is an abstract version of the Gödel incompleteness theorem, in the sense that Gödel's construction provides

a constructive method that given a formal system for arithmetic in which only true statements are provable, produces another statement (the Gödel sentence for the system) that is also true, but unprovable in the system.

If S is a creative set, then its complement $\overline{S}$ contains an infinite recursively enumerable subset. This follows from the fact that starting from the empty set, we can construct successively larger subsets of $\overline{S}$ using the function $f(i)$ to obtain an infinite recursively enumerable subset of $\overline{S}$. If A and B are recursively enumerable, A is creative, and A is many-one reducible to B, then B is also creative. It follows that if B is an m-complete recursively enumerable set, that $\overline{B}$ contains an infinite recursively enumerable subset.

This last result motivates Post's next definition: a recursively enumerable set S of positive integers is *simple* if $\overline{S}$ is infinite, but does not contain any infinite recursively enumerable subset. A simple set cannot be recursive, for if it were, $\overline{S}$ would be an infinite recursively enumerable subset of itself. Post's next result is a clever diagonalization that proves that a simple set exists. For every basis B_i, his construction guarantees that a member of W_i (the set generated by B_i) is placed in S, provided W_i is infinite, so that this member is a witness to the fact that W_i is not a subset of $\overline{S}$. At the same time, Post constrains the number of witnesses that can be added so that $\overline{S}$ is infinite.

The existence of a simple set shows that there is a recursively enumerable set of intermediate degree in the ordering according to many-one reducibility. In fact, this result can be strengthened by defining a more general notion of reducibility. If a set A is reducible to a set B, then we can think of the reduction as being given in terms of membership queries. So, for example, if A is many-one reducible to B, we put an integer n in A if the answer to the query "Is $f(n)$ in B?" is positive, otherwise we leave it out. However, we could just as well leave n out if the answer is positive, and otherwise put it in. In other words, it is quite reasonable to say that the complement $\overline{S}$ of a set is reducible to the set S itself. This idea leads immediately to the more general notion of truth-table reducibility.

Define a *Boolean query to a set* B to be a propositional formula built out of the Boolean connectives $\wedge, \vee, \sim$, in which the atomic formulas are all of the form $(m \in B)$, where m is a positive integer; for example, the formula

$$[(5 \in B) \wedge \sim(67 \in B)] \vee [\sim(42 \in B) \wedge (87 \in B)].$$

We define a set A to be *truth-table reducible* to a set B if there is a recursive function f so that for all n, $f(n)$ is a Boolean query to B, and n is in A if and only if B satisfies the query $f(n)$. If, in addition, there is a fixed upper bound on the number of atomic formulas in the Boolean queries $f(n)$, then we say that A is *bounded truth-table reducible* to B.

Post is able to show that if S is a simple set, then a creative set cannot be reduced to S by bounded truth-table reductions. However, this result does not generalize to unbounded truth-table reductions, since he proves that if C is any creative set, then there is a simple set S so that C is reducible to S by unbounded truth-table reductions.

This last result leads Post to generalize the notion of simplicity to that of a *hyper-simple set*. A hyper-simple set H is defined to be a recursively enumerable set of positive integers so that $\overline{H}$ is infinite, but there is no infinite recursively enumerable set of mutually exclusive finite sequence of positive integers so that each sequence has at least one member in $\overline{H}$ ("mutually exclusive" means that the sequences have no elements in common). The notion of hyper-simplicity is powerful enough to show that there are intermediate degrees of unsolvability in the truth-table ordering, since Post proves, first that a hyper-simple set exists, and second that no creative set can be reduced to a hyper-simple set by a truth-table reduction.

Although the notion of truth-table reducibility is fairly broad, it is far from the most general notion of reducibility that can be defined. Let us imagine that we are trying to decide whether or not an integer belongs to a set A, while having access to information about the answers to all possible queries of the form: "Is integer n in set B?" Alan Turing, in his doctoral thesis [Turing, 1939, §4], [Davis, 1965, p. 166] suggested the picturesque metaphor of having access to an "oracle" for the set B, a terminology that is now standard. Then we say that A is *Turing-reducible* to B if we can decide whether or not an integer is in A, using a computer that has access to an oracle for B.

The definition given in the previous paragraph is not mathematically precise, but it can be made so in a variety of ways. The method usually adopted is to define a Turing machine with an auxiliary tape on which numbers can be written as queries to the oracle. When the machine enters a query of the form "Is the integer n in the set B?", the computation continues in one of two different ways, depending on the answer from the oracle (the oracle's answers are considered to be instantaneous, so that a query counts as a single computational step). Post himself described an alternative approach [Post, 1948] based on a generalization of his canonical sets to S-canonical sets, in which primitive assertions representing propositions of the form "Integer n is in S" and "Integer n is not in S" are added to a canonical system.

Post was unable to solve the problem of whether there exist intermediate degrees of unsolvability for the most general notion of reducibility, Turing-reducibility. This became known as "Post's problem," and work towards its solution drove a great deal of the research stimulated by his paper. He expressed some doubts as to whether hyper-simple sets would provide examples of intermediate degrees, but was not able to decide the question. The matter was not fully clarified until after Post's problem itself had been solved. The paper [Jockusch and Soare, 1973] shows that if H is Post's original example of a hyper-simple set, then it may or may not be Turing-complete, depending on the precise enumeration of bases $B_1, B_2, \ldots, B_n, \ldots$ adopted.

Post's program of research aimed at solving the problem of the existence of intermediate degrees was to define recursively enumerable sets with "sparse" complements, an idea implemented in the definition of simple and hypersimple sets. However, although this approach succeeded in the case of the strong reducibilities,

he was not able to make it work for Turing reducibility. The solution to the problem was finally achieved by a different approach, the priority method of Friedberg and Muchnik, discussed below in §6.5.

6.3 The Post Correspondence Problem

In his brief note [Post, 1946], Post provided an addendum to [Post, 1943], in which he presents an unsolvable problem of striking simplicity. He considers strings composed of the two symbol alphabet $\{a, b\}$. An instance of the *correspondence problem* is given by a finite list $(g_1, h_1), (g_2, h_2), \ldots, (g_m, h_m)$ of pairs of non-empty strings, for example, the list $(bb, b), (ab, ba), (b, bb)$. A *solution* to such a problem consists of a finite sequence $\sigma = \sigma_1, \ldots, \sigma_k$, where $1 \leq \sigma_i \leq k$, so that $g_{\sigma_1} g_{\sigma_2} \ldots g_{\sigma_k} = h_{\sigma_1} h_{\sigma_2} \ldots h_{\sigma_k}$; a solution to the example just given is provided by the sequence 1223, since in this case $g_1 g_2 g_2 g_3 = bbababb = h_1 h_2 h_2 h_3$.

The preceding example has a very simple solution. Much more difficult instances of the correspondence problem can be constructed. As a challenge to the reader, here are two instances that are known to be solvable, but whose solutions are very difficult:

Instance 1: [Lorentz, 2001] $(abb, a), (b, abb), (a, b)$;

Instance 2: [Zhao, 2002] $(aaba, a), (baab, aa), (a, aab)$.

Post proves unsolvability of his correspondence problem by reducing the decision problem for normal systems to it. He does not give any applications of the problem. However, because of its simplicity, it later turned out to be an ideal problem for proving undecidability results in formal language theory. Bar-Hillel, Perles and Shamir [Bar-Hillel *et al.*, 1961] applied the correspondence problem to show that several problems in the theory of formal languages are undecidable. Among the problems they prove undecidable is that of deciding whether two context-free languages have a non-empty intersection (that is, given two phrase structure grammars, determining whether or not the languages generated by the grammars have a word in common). The Post correspondence problem is now a standard tool of formal language theory – see for example the textbook [Hopcroft and Ullman, 1979].

6.4 The Word Problem for Semigroups

By the 1940s, the broad significance of the concept of recursive, or computable functions was already clear. The universality and robustness of the concept of computability, and its applicability to the classical decision problems in logic was well known to the logical community. Nevertheless, the applications had remained within the realm of logic, and the fruitfulness of the idea outside this area remained to be shown. Post provided a major step in this direction by demonstrating that a decision problem that had arisen outside the area of logic and foundations was unsolvable.

The Norwegian mathematician Axel Thue in his paper [Thue, 1914] had posed the word problem for finitely presented semigroups. The word problem for such a semigroup is given as follows. There is a finite set of primitive letters, representing the generators of the semigroup, from which words can be formed using a binary associative operation. For example, if the generators of the semigroup are a, b, c, then the words $((ab)c)a$ and $a(b(ca))$ are considered to be the same, so that the words are treated as strings of letters. Then a *presentation* of a semigroup with these generators is given by a finite set of equations, for example, the equations $ab = ba$ and $abc = bcaa$. The word problem is, given such a presentation, to determine whether or not a given equation is derivable using the standard rules for reasoning with equations, where the generators are treated as constants.

Alonzo Church suggested to Post that Thue's problem, might be proved unsolvable as an application of [Post, 1946]. In fact, however, Post solved the problem not by using his correspondence problem, but by translating the formalism of Turing machines into the notation of his canonical systems. A presentation of a semigroup can be rewritten as a set of production rules of the form $P\sigma Q \Longrightarrow P\tau Q$, where σ and τ are strings over a finite alphabet. A set of production rules of this form Post defines as a *semi-Thue system*. A *Thue system*, which corresponds to a semigroup presentation, has the added property that whenever the production $P\sigma Q \Longrightarrow P\tau Q$ is in the set, so is its inverse $P\tau Q \Longrightarrow P\sigma Q$. For example, the presentation given above can be rewritten as the set of productions

$$PabQ \Longrightarrow PbaQ, PbaQ \Longrightarrow PabQ, PabcQ \Longrightarrow PbcaaQ, PbcaaQ \Longrightarrow PabcQ.$$

The decision problem for a semi-Thue system is the problem of determining whether or not a word τ can be derived from a given word σ using the production rules of the system. The word problem of Thue is equivalent to the decision problem for Thue systems. Post's strategy to prove unsolvability of this problem is to reduce the decision problem for a special class of semi-Thue systems to the decision problem for Thue systems.

This proof strategy is most easily carried out for rules that are *deterministic*, in the sense that at most one rule is applicable to any string derivable from an initial assertion. This is why canonical systems derived from Turing machines are ideal for this problem. A configuration of a Turing machine can be described as a string consisting of an endmarker h, followed by a sequence of tape symbols, followed by a symbol for an internal state of the machine, then another string, and finally another endmarker symbol h. For example, the string $h110q_301h$ represents a configuration of a machine in which the internal state of the machine is q_3, and the machine is scanning the fourth symbol of the string 11001 on an otherwise blank tape. The rules of the Turing machine are then easily rewritten as Post production rules, for example, the rule $Pq_31Q \Longrightarrow P1q_3Q$ represents the Turing machine instruction "In state q_3, if you are scanning a 1, then move right."

Post starts from an unsolvable problem in the theory of Turing machines. It is an unsolvable problem to determine whether or not a Turing machine, with internal states $q_1, \ldots, q_r$, started in its initial state q_1, reading the first symbol in

a binary string σ on its input tape, halts in its highest-numbered state q_r on a blank tape. Rewriting the machine instructions as a semi-Thue system T', this decision problem becomes: Is there a derivation in the system T' of the string hq_rh from the initial string $hq_1\sigma h$? Let T'' be the system consisting of all the inverses of the productions in T'. Then hq_rh is derivable in T' from $hq_1\sigma h$ if and only if $hq_1\sigma h$ is derivable from hq_rh in T'. Now let T be the system that has all the rules of T', together with all the rules of T'' (that is, all the rules of T' and their inverses). Then because of the deterministic character of Turing machine rules, exactly the same strings are derivable from the initial string hq_rh in T as are derivable in T''. This reduces the Turing machine problem to the decision problem for a Thue system, and so shows Thue's problem of 1914 to be unsolvable.

Very simple unsolvable cases of Thue's problem are known. Building on Novikov's proof of the unsolvability of the word problem for groups (see below), [Tseitin, 1958] shows that the word problem for the semigroup with five generators $\{a,b,c,d,e\}$ and defining relations $ac = ca, ad = da, bc = cb, bd = db, eca = ce, edb = de, cca = ccae$ is unsolvable.

The formulation of Turing machines used by Post is not the original formulation of Turing, but a simplified and clarified version due to Post himself. It is in fact Post's version of the Turing machine that is the standard "Turing machine" presented in introductory texts on logic and computer science. Post includes an appendix to his paper providing a detailed critique of Turing's original paper.

The Russian logician Markov proved the same result independently and almost at the same time as Post [Markov, 1947]. These results of Post and Markov were just the beginning of a host of mathematical problems proved undecidable by the technique of reducing known unsolvable problems to them. A notable result achieved somewhat later is the unsolvability of the word problem for finitely presented groups, proved by the Russian logician P.S. Novikov in [Novikov, 1955], and independently though somewhat later by the American logician William Boone [Boone, 1957a][Boone, 1957b].

A few years later, Post showed that another concrete decision problem (though of a logical nature) was undecidable, in joint work with Samuel Linial (who later changed his name to Samuel Gulden). Linial and Post showed that there is an undecidable subsystem of classical propositional logic by showing how to encode a system of normal productions in a propositional calculus with detachment as a primitive rule. This research was published only as an abstract [Linial and Post, 1949]; an exposition of their results is to be found in [Davis, 1958, pp. 137-142].

6.5 Degrees of unsolvability

The joint paper [Kleene and Post, 1954] is an important landmark in the development of recursion theory. It had its origins in an encounter between Kleene and Post at the meeting of the American Mathematical Society at which Post presented [Post, 1944] on February 26 1944. In an interview of 1984, Kleene recalled:

> I was in the Navy when he presented his paper "Recursively Enumer-

> able Sets ..." at a meeting of the American Mathematical Society. I was teaching midshipmen in the U.S.S. John Jay. That's not a ship, it's a Columbia dormitory. But when midshipmen entered, they said "Request your permission to come aboard, sir." I only had to walk 200 yards to go to the lecture. After the lecture I had Post over to my apartment. I had an apartment just off the Columbia campus. There I presented to him – I think it was already in press, if it wasn't, it was in manuscript – my paper, "Recursive Predicates and Quantifiers," which had a very close relation to what he was doing. That's the first he knew of that [Aspray, 1984].

This paper [Kleene, 1943] is a fundamental contribution in which the arithmetical hierarchy is introduced. The bottom level of Kleene's hierarchy is the family of recursive sets of natural numbers, denoted by $\Sigma_0 = \Pi_0$. For $n > 0$, a set B of natural numbers is in Σ_n if there is a recursive relation $R(x, y_1, \ldots, y_n)$ so that for any natural number x,

$$x \in B \iff \exists y_1 \forall y_2 \exists y_3 \ldots Q y_n R(x, y_1, \ldots, y_n),$$

where Q is $\exists$ if n is odd, and $\forall$ if n is even. Similarly, a set B of natural numbers is in Π_n if there is a recursive relation $R(x, y_1, \ldots, y_n)$ so that for any natural number x,

$$x \in B \iff \forall y_1 \exists y_2 \forall y_3 \ldots Q y_n R(x, y_1, \ldots, y_n),$$

where Q is $\exists$ or $\forall$ according as n is even or odd. This hierarchy of sets generalizes the classification of sets as recursive and recursively enumerable, since the family of recursively enumerable sets is identical with Σ_1, while the complements of such sets (co-r.e. sets) constitute Π_1. By generalizing the diagonal argument that shows the existence of recursively enumerable but non-recursive sets, Kleene proves that this hierarchy is proper. That is to say, the obvious inclusions $\Sigma_n \subseteq \Sigma_{n+1}$, $\Pi_n \subseteq \Pi_{n+1}$, $\Sigma_n \subseteq \Pi_{n+1}$ and $\Pi_n \subseteq \Sigma_{n+1}$ are all proper.

Post was stimulated by Kleene's work to generalize his work on reducibilities between recursively enumerable sets to a much broader theory of degrees of unsolvability in work reported in the abstract [Post, 1948]. Recall from §4.1 Post's notion of "finite-normal-test" – a set M of integers is decidable just in case there is a finite-normal-test for it, that is, a normal system that has assertions exactly representing which integers are and are not elements of M. Post generalizes his earlier notion of canonical sets to S-canonical sets by adding to his canonical systems primitive assertions representing membership and non-membership of integers in S, where S is an arbitrary set of integers. Using this broader concept, he generalizes his earlier characterization of decidable sets by showing that a set S_1 is Turing-reducible to a set S if and only if both S_1 and its complements are S-canonical sets.

Furthermore, the existence of the complete set K generalizes to the relativized world of oracle computations. If A is an arbitrary set of integers, then there is a set $A' = K^A$ that has the property: K^A can be enumerated by a machine employing

an oracle for A, and furthermore, any set of integers that can be enumerated by a machine employing an oracle for A (an A-canonical set) is reducible to K^A. If $\mathbf{a}$ is the Turing degree of the set A, define $\mathbf{a}'$ (the *jump* of $\mathbf{a}$) to be the degree of the set $A' = K^A$. As a generalization of the theorem proving the existence of recursively enumerable sets that are not recursive, we can show that $\mathbf{a}$ is reducible to $\mathbf{a}'$, but $\mathbf{a}'$ is not reducible to $\mathbf{a}$.

This definition results in a hierarchy of degrees of unsolvability, with respect to Turing reducibility. If $\mathbf{0}$ is the degree of $\emptyset$ (the degree of the recursive sets), then $\mathbf{0}'$ is the degree of the complete set K, and this is the beginning of a strictly ascending chain of degrees $\mathbf{0}, \mathbf{0}', \mathbf{0}'', \ldots$ that can be extended to transfinite levels. The results of Post just quoted imply a relationship between the arithmetical hierarchy of Kleene and Post's hierarchy of degrees, a result usually called *Post's Theorem*. Post's theorem says that $\Sigma_{n+1} \cap \Pi_{n+1}$ is exactly the sets of integers reducible to $\mathbf{0}^{(n)}$, where $\mathbf{0}^{(0)} = \mathbf{0}$, and $\mathbf{0}^{(n+1)} = (\mathbf{0}^{(n)})'$. Hence, Post's degree hierarchy is interleaved between the levels of Kleene's hierarchy.

Post's theorem is the main completed result reported in the abstract [Post, 1948]. However, he hints at further results. "Work is in progress on further equivalence proofs, further applications of the K_n-scale, incomparable degrees of unsolvability related to the K_n-scale, extension of the main theorem to Mostowski, Fund. Math. vol. 34, and extension of the K_n-scale into the constructive transfinite." These later results, though, did not appear in print, and Kleene had occasion to remonstrate with Post on this score. Kleene tells the story in a conversation (also involving Andrzej Mostowski, Gerald Sacks, Michael Morley and Anil Nerode) recorded in [Crossley, 1975, pp. 18-20].

> **Mostowski:** I must raise a protest against this habit of not publishing. That is all right here, because you meet every second month to collaborate at a conference here or there, but people like me are completely cut off.
>
> **Kleene:** Yes.
>
> **Sacks:** I agree completely.
>
> **Morley:** It should lapse.
>
> **Sacks:** And anyone who does not publish his work should be penalized.
>
> **Kleene:** This is just what I wrote to Emil Post, on construction of incomparable degrees and things like that, and he made some remarks and hinted at having some results and I said (in substance): "Well, when you leave it this way, you say you have these results, you don't publish them. The fact that you have them prevents anyone else who has heard of them from doing anything on it." So he said (in substance): "You have sort of pricked my conscience and I shall write something out," and he wrote some things out, in a very disorganized form, and he suggested that I give them to a graduate student to turn into a paper. As I recall, I think I did try them on a graduate student, and

the graduate student did not succeed in turning them into a paper, and then I got interested in them myself, and the result was eventually the Post-Kleene paper [Kleene and Post, 1954].

Morley: You mean one of your graduate students could have had the Post-X paper?

Nerode: You mean one that wanted to work hard.

Kleene: I suppose, maybe it was not good for a graduate student, because a graduate student needs a thesis he can publish under his own name, and this would have had to be joint, or maybe ...

Morley: Oh, I don't know. I think he could have borne having a Post-X paper.

Kleene: As a matter of fact, it could have – if a graduate student had picked it up. There were things that Post did not know, like that there was no least upper bound. [1] You see, Post did not know whether it was an upper semi-lattice or a lattice. I was the one who settled that thing.

Sacks: What are you talking about? The degrees of the arithmetic sets?

Kleene: No. The upper semi-lattice of degrees of unsolvability.

The paper [Kleene and Post, 1954] was written by Kleene, and its precise but very formal style forms a stark contrast with Post's own rather discursive mathematical prose. In it, the degree hierarchy is defined in terms of Kleene's own formalism of general recursive functions [Kleene, 1936], [Kleene, 1952a, Chapter XI]. The notion of reducibility is that of a function being general recursive in certain other functions, to which various earlier notions of reducibility were known to be equivalent. Post had proved in unpublished work [Post, 1948] that his own notion of canonical reducibility was equivalent to Turing reducibility [Turing, 1939], Martin Davis had proved in his Princeton doctoral thesis [Davis, 1950] that canonical reducibility is equivalent to general recursive reducibility, while Kleene had given a direct proof of the equivalence of Turing reducibility and general recursive reducibility [Kleene, 1952a, §68, §69].

The relation "A is recursive in B" is reflexive and transitive, so the relation "A is recursive in B and B is recursive in A" is an equivalence relation. Hence, this equivalence relation partitions the family of all sets of natural numbers into equivalence classes, or *degrees of unsolvability*. There are $\aleph_0$ sets in each equivalence class and hence $2^{\aleph_0}$ degrees, since there are $2^{\aleph_0}$ sets of natural numbers. The family of degrees forms an upper semi-lattice, that is, given degrees $\mathbf{a}$ and $\mathbf{b}$, there is a least upper bound $\mathbf{a} \cup \mathbf{b}$ in the degree ordering. This is because, given any two sets A and B of degrees $\mathbf{a}$ and $\mathbf{b}$, we can define a set C (for example,

[1] There seems to be an error in the transcript here. Presumably "greatest lower bound" is meant.

$\{2a : a \in A\} \cup \{2b+1 : b \in B\}$) so that A and B are both recursive in C, while C is recursive in A, B.

The main result for which [Kleene and Post, 1954] is remembered is the proof of the existence of incomparable degrees between a degree $\mathbf{a}$ and its jump $\mathbf{a}'$. The technique employed by Kleene and Post involves a countable set of conditions, or "requirements," as they are usually called today. Given a set B, there are a countable number of Turing machines $M_0, M_1, \ldots, M_k, \ldots$ that use an oracle for B. Hence, we can give an effective list of requirements of the form: "The set C is not identical with the set computed by Turing machine M_k with an oracle for B," and a similar list with C and B interchanged. Any such condition can be fulfilled by ensuring the presence or absence of a particular integer in B or C, so Kleene and Post's construction consists in constructing B and C step by step, while ensuring that all such conditions are eventually fulfilled. Their construction of A and B, however, requires an oracle for the set K^A (where A has the degree $\mathbf{a}$) and so it does not provide an answer to Post's problem.

The paper contains a rich collection of additional results. The construction of intermediate degrees sketched in the preceding paragraph is generalized to show that between the degrees $\mathbf{a}$ and $\mathbf{a}'$, there is an infinite linearly ordered sequence of degrees that is dense, that is to say, if $\mathbf{c}, \mathbf{d}$ are degrees in the sequence, where $\mathbf{c} < \mathbf{d}$, there is a degree $\mathbf{e}$ so that $\mathbf{c} < \mathbf{e} < \mathbf{d}$. The paper also shows that the degree ordering is not a lattice, by showing that there are degrees $\mathbf{a}$ and $\mathbf{b}$ with no greatest lower bound in the degree ordering (this is the result to which Kleene alludes in the transcript above). Although the paper made considerable inroads into determining the fine structure of degrees, it also left many open problems, and provided a powerful stimulus for further work by later logicians.

It also provided an essential stepping stone in the solution of Post's problem by two young mathematicians, the first an American [Friedberg, 1957] and the second a Russian [Muchnik, 1956]. Both Friedberg and Muchnik followed Kleene and Post in constructing a pair of incomparable degrees by the technique of listing an infinite set of requirements that are all fulfilled at the end of the construction. To avoid the use of an oracle, however, requires a more subtle technique. In the original Kleene-Post method, a requirement, once fulfilled, remains fulfilled permanently and requires no further attention. In the constructions of Friedberg and Muchnik, however, a requirement can be violated, or "injured," at an intermediate stage. To ensure that a requirement is eventually fulfilled in spite of these injuries, Friedberg and Muchnik impose a priority ordering on their requirements, ensuring that only finitely many injuries can happen to a given requirement, so that it is eventually satisfied. Consequently, this new technique, which is one of the fundamental methods in recursion theory, is now known as the "priority method."

Since the original solution of Post's problem, recursion theory has become a discipline of extraordinary depth and richness, and the priority method has been used in more and more complex settings. Excellent expositions of the resulting theory are available in the books [Rogers, 1967], [Shoenfield, 1971], [Soare, 1987], [Odifreddi, 1989] and [Odifreddi, 1999].

7 PROVABILITY AND DEFINABILITY

The last major project of Post's career grew out of his early interest in the question of absolutely undecidable propositions. Logical developments in the early twentieth century had apparently shown that notions of provability and definability are not absolute. In the case of provability, the incompleteness theorems demonstrate that any consistent formal axiomatic theory for elementary number theory can be consistently extended by the addition of an unprovable sentence, so that it seems that we can not identify any fixed or absolute notion of provability. Similarly, it would appear that the diagonal method, applied to any constructively defined collection of definable notions, allows us to find a new definable concept not included in the original list.

Post was very impressed by the absolute character of the concept of computable function, since the extent of the computable functions does not depend on the formalism chosen, a point that he emphasized repeatedly. This led him to hope that a similar absolute notion of provability might be attainable, leading to absolutely undecidable propositions.

Already in the early 1920s, Post was interested in the question of the existence of propositions that are absolutely undecidable in some sense. He describes "A Probably Fallacious Suggestion for a Non-Provable Theorem" in the Appendix to his unpublished manuscript of 1941 [Post, 1965, pp. 421-422] [Post, 1994, p. 432]. He considers the enunciation: "For each deductive system of the normal form and enunciation in it there exists a finite method of proving or disproving the derivability of the enunciation in the system," and attempts to argue that if it were provable or disprovable, this would contradict the unsolvability of the decision problem for normal systems.

The preceding attempt proved abortive; Tarski pointed out to Post that his attempted argument was invalid because the supposed proof might be non-constructive. However, he continued his attempts along these lines, and in a footnote to the 1941 manuscript [Post, 1965, pp. 341-342][Post, 1994, p. 376] says that "since February 1938 we have given an occasional week to a continuation of this work, and largely in the spirit of the Appendix. Our goal, however, is now an analysis of proof, perhaps leading to an absolutely undecidable proposition, rather than an analysis of finite process."

The question of the existence of absolutely undecidable propositions was one of the topics that he discussed with Gödel at their meeting in October 1938, as appears from an entry of 4 November 1940 in one of Post's notebooks:

> In first meeting with Gödel (about a year ago or more [probably spring 1939]) [2] suggested to him abs. undec. prob. He said perhaps Cont. Hyp. I said this more like parallel axiom, i.e. would merely mean different theories of classes. Wanted rather abs. undec. arith. prop. where analyse all possible methods of proof & perhaps find a property

[2] In fact the meeting was on October 29 1938, as described above.

> of all such which would lead to abs. und. arith. prob. Gödel poo-poohed idea. Said (roughly) it was absurd [Grattan-Guinness, 1990, pp. 82-83].

It appears that Post continued this discussion with Gödel at a second meeting, for his diary entry continues:

> In second meeting, Oct. 1940, I asked him what he was working on now ... Said to prove negative of con. hyp. consistent with rest of set theory. Said he hadn't succeeded yet, but if so would be abs. undec. prob. Raised ‖ axiom analogy again. But he said no, that axioms of set theory categorical for all models [Grattan-Guinness, 1990, p. 83].

Gödel expressed an opinion in print around this time related to those reported by Post (though considerably more cautious). In his first announcement of his consistency proof [Gödel, 1938], he wrote about the axiom of constructibility ("proposition A"), from which the generalized continuum hypothesis can be deduced within the system T of set theory without the axiom of choice:

> The proposition A added as a new axiom seems to give a natural completion of the axioms of set theory, in so far as it determines the vague notion of an arbitrary infinite set in a definite way. In this connection it is important that the consistency proof for A does not break down if stronger axioms of infinity (e.g., the existence of inaccessible numbers) are adjoined to T. Hence the consistency of A seems to be absolute in some sense, although it is not possible in the present state of affairs to give a precise meaning to this phrase.

Later, of course, he changed his mind on this point, and in [Gödel, 1947] espoused the view that new axioms of set theory would be found that would decide the continuum hypothesis (most likely in a negative sense).

Post, however, was not looking for absolutely undecidable propositions in the realm of set theory, where he appears to have thought that alternative set theories were possible, in analogy with alternative theories of geometry. His own goal was to find such problems in the realm of arithmetic. In his paper of 1943, he refers again to the problem in a footnote to a passage in which he states that the decision problem for normal systems is unsolvable:

> Absolutely unsolvable, that is, to use a phrase due to Church. By contrast, the undecidable propositions of Gödel's epoch making paper of 1931 ... are but relatively undecidable, the very proof of their undecidability in the given logic leading to an extension of that logic in which they are, indeed, proved to be true. A fundamental problem is the question of the existence of absolutely undecidable propositions, that is, propositions which in some *a priori* fashion can be said to have a determined truth-value, and yet cannot be proved or disproved by any valid logic [Post, 1943, p. 200], [Post, 1994, p. 445].

Even though Post reported that Gödel had "poo-poohed" his idea of absolutely undecidable propositions in 1938, it is significant that in his Gibbs lecture of 1951, Gödel asserted the possibility of absolutely undecidable arithmetical propositions, though only as one half of a disjunction. The lecture is devoted to the philosophical consequences of the incompleteness theorems. Gödel considers arguments somewhat resembling those given in [Lucas, 1961] and [Penrose, 1989], but (unlike these later writers) drawing not an absolute but a disjunctive conclusion:

> Either mathematics is incompletable in this sense, that its evident axioms can never be comprised in a finite rule, that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems of the type specified [Gödel, 1995, p. 310].

Post continued to work along the lines suggested by the footnote to his 1943 paper, searching for a concept of provability in arithmetic that could play a similar role in the theory of numbers as recursivity in the theory of algorithms. The only remaining published record of his thought on this subject in the late 1940s is in the brief abstract [Post, 1953b], in which he says:

> In 1947, in a lost letter to Tarski via Church, the author proposed a formulation of "Primitive Inductive-Reflective Proof" involving, besides the elementary, only mathematical induction and "Gödelization." But representation theory for Gödel's system P (1931) failed to materialize due to P's Axiom of Reducibility. Indeed, Kleene's Example (Proceedings of the International Congress of Mathematicians, 1950, vol. II, p. 683) offers hope of an impossibility proof.

In his Princeton doctoral thesis [Turing, 1939] [Feferman, 1995] Turing studied the possibility of overcoming arithmetical incompleteness by the successive adjunction of unprovable sentences, given by Gödel's construction – this is presumably the process that Post describes as "Gödelization." Post seems to have hoped to show that the adjunction of axioms involving higher types was in some sense conservative with respect to first-order arithmetical sentences, but encountered insuperable difficulties in connection with the type-theoretical comprehension axiom, or "Axiom of Reducibility," as he calls it. The result of Kleene to which Post alludes is a theorem in [Kleene, 1952b] showing that Brouwer's fan theorem [Brouwer, 1924] fails if "choice sequence" is interpreted to mean "recursive function." In classical terms, he shows that there is an infinite recursive binary tree with no infinite recursive branch [Odifreddi, 1989, p. 506].

The conclusion that Post drew from his abortive attempts towards constructing an absolute notion of provability is that he was attacking the problem in the wrong order. The right order, he now thought, was: solvability, definability, provability. The immediately succeeding abstract [Post, 1953a] reports his success in finding a candidate for the role of an absolute notion of definability.

An obvious difficulty that arises in attempting to give an absolute notion of definability is this: if not all ordinals are definable, then there must be a least indefinable ordinal (this will certainly be true if we have only countably many definitions). But if our notion of definability is itself definable, then a contradiction (Richard's paradox) results immediately, since the least indefinable ordinal would then be definable.

A possible way out of this impasse is to take the notion of definability to be highly non-constructive, so that the definable sets form a proper class, and all ordinal numbers are definable. Post carries out this idea by extending the simple theory of types of order ω, as formulated in [Gödel, 1931], to transfinite levels, with a type level for all ordinals α. Thus Gödel's system P is extended to a system $P^{(\alpha)}$, with variables $x_0^{(\alpha)}, x_1^{(\alpha)}, \ldots$ ranging over sets of type α in the type hierarchy, and the usual apparatus of connectives and quantifiers. Post remarks: "The present thesis is that every definable set is given by some α-formula" (his intention was to give a necessary condition for definability, rather than a full-fledged explication, as in the case of computability).

Post was unaware that a closely related absolute concept of definability had been proposed earlier by Gödel, in his remarks at the Princeton Bicentennial Conference on Problems in Mathematics held in 1946. Gödel's paper was not published at the time, and the full text of his address first appeared in print in the collection edited by Martin Davis [Gödel, 1965, pp. 84-88].

Gödel's remarks strikingly parallel Post's own thinking on these topics. Commenting on an earlier lecture by Tarski, he remarks that the great importance of the concept of general recursiveness is largely due to the fact that "with this concept one has for the first time succeeded in giving an absolute definition of an interesting epistemological notion, i.e., one not depending on the formalism chosen." Contrasting the situation with that for provability and definability, he says:

> By a kind of miracle it is not necessary to distinguish orders, and the diagonal procedure does not lead outside the defined notion. This, I think, should encourage one to expect the same thing to be possible also in other cases (such as demonstrability or definability) [Gödel, 1965, p. 84].

Gödel goes on to make a very tentative suggestion about an absolute concept of provability, conjecturing that any set-theoretical assertion can be decided by a sufficiently powerful axiom of infinity, so that "every proposition expressible in set theory is decidable from the present axioms plus some true assertion about the largeness of the universe of all sets."

He continues with "somewhat more definite suggestions," proposing definability in terms of ordinals as an absolute notion of definability in set theory. His proposal is to extend the usual language of Zermelo-Fraenkel set theory by including primitive names for all ordinal numbers (so that the formulas of the extended language form a proper class); ordinal definable sets are those specified by a formula

from this extended language. This concept of definable set is broader than his earlier notion of constructible set, used in proving the consistency of the generalized continuum hypothesis with the axioms of set theory, since the quantifiers in the extended language range over all sets, whereas constructible sets are defined by quantifiers restricted to the constructible universe.

Post's own notion is slightly more restrictive than that of Gödel, and is now known as that of "hereditary ordinal definability"; a set is hereditarily ordinal definable if it is not only ordinal definable, but its members, members of members ...etc. are also ordinal definable. The notion of ordinal definability was rediscovered again independently during the 1960s [Takeuti, 1961], [Myhill and Scott, 1971].

Because of Post's untimely death, there is no other published record of his last logical project. However, two notebooks on "Definability" written in 1952-1954 remain in the archives of the American Philosophical Society in Philadelphia, and may still repay investigation.

8 POST'S LEGACY

The remarkable fertility of Post's ideas, particularly in the areas of recursion theory and computer science, is attested by the number of concepts and results named after him: "Post-completeness," "Post's problem," "Post productions," "Post correspondence problem," "Post's classification theorem" are some of the more prominent examples. His influence on computer science is most remarkable, considering that he seemingly never had any contact with computing machinery or computer programming.

Post's ideas also bore fruit in the area of formal linguistics. The early work of Noam Chomsky on phrase structure grammars was influenced by the ideas of Post; Chomsky formulates basic notions such as that of a context-free language in terms of rules closely akin to Post productions [Chomsky, 1956]. He seems to have known of these ideas indirectly through the unusual textbook of Paul Rosenbloom [Rosenbloom, 1950], in which Post's production systems are given a starring role, particularly in the last chapter. Seymour Ginsburg, an important early contributor to formal language theory, and author of a well known monograph giving a comprehensive overview of early work on context-free languages [Ginsburg, 1966], was a student with Martin Davis in an honors mathematics class taught by Post [Gleiser, 1980]. All five of the students in this course, described by Davis as a "pressure cooker experience," became mathematicians.

Post also exerted influence through the students who were inspired by him and his remarkable classes. Martin Davis has described Post's teaching methods:

> Post's classes were tautly organized tense affairs. Each period would begin with student recitations covering problems and proofs of theorems from the day's assignment. These were handed out apparently at random and had to be put on the blackboard without the aid of

> textbook or notes. Woe betide the hapless student who was unprepared. He (or rarely she) would have to face Post's "more in sorrow than anger" look. In turn, the students would recite on their work. Afterwards, Post would get out his 3 by 5 cards and explain various fine points. The class would be a success if he completed his last card just as the bell rang. Questions from the class were discouraged: there was no time. Surprisingly, these inelastic pedagogic methods were extremely successful, and Post was a very popular teacher [Davis, 1994, p. xxv].

David Hilbert, in his address before the International Congress of Mathematicians in Paris in 1900, had posed a series of problems to challenge mathematicians in the new century. Hilbert's tenth problem reads as follows:

> Given a diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: *To devise a process according to which it can be determined by a finite number of operations whether the equation is solvable in rational integers* [Hilbert, 1902].

A diophantine equation is any algebraic polynomial equation with whole number coefficients, for example: $31x^2 + 41y^3 - 17x = 0$. Hilbert's problem asks for an algorithm to determine whether or not an arbitrary diophantine equation is solvable (in line with his rationalistic and optimistic outlook, he seems to have expected a positive solution). In his address of 1944, Post remarked that Hilbert's problem "begs for an unsolvability proof" [Post, 1944, p. 289]. The challenge was taken up by Post's student Martin Davis, who was one of the key contributors to the proof of unsolvability, together with Hilary Putnam and Julia Robinson. The final step completing the proof was provided by the young Russian mathematician Yuri Matiyasevich [Matiyasevich, 1970]. The reader will find excellent and very readable accounts of the history of this problem in [Reid, 1996] and [Yandell, 2002]. A fully detailed account of the proof, together with much else, is in the beautifully written book [Matiyasevich, 1993].

8.1 Post's Papers

Post's papers are housed at the American Philosophical Society in Philadelphia. They include correspondence, unpublished manuscripts, research notes, and photographs documenting Post's career. The notebooks were donated by Martin Davis in 1986; the remaining papers were donated by Phyllis Post Goodman in 1992. The correspondence includes letters from Paul Bernays, Alonzo Church, Martin Davis, Frederic Fitch, Abraham Fraenkel, Kurt Gödel, Stephen Kleene, W.V. Quine and Alfred Tarski. Ivor Grattan-Guinness has given a description of the papers [Grattan-Guinness, 1990], and has transcribed an unpublished manuscript from 1935 [Post, 1990]. This manuscript is a somewhat sketchy essay on Russell's paradox, in which Post works out the idea that the class of existing classes may not be

fixed, but rather can expand through the process of thinking about existing objects. This vision of mathematical objects as processes that evolve through time is of course one of the fundamental ideas that runs through all of Post's work, starting from his dissertation, where he introduces the class of elementary propositions by describing "the vision of the totality of these functions streaming out from the unmodified variable p through forms of ever growing complexity ..." [Post, 1921a, p. 165].

This obsession with time also manifests itself throughout the surviving notebooks. There are four major numbered series, five volumes on "Closed Truth Systems: Towards a new presentation," written in 1929-1931 when Post was working out the new version of his results on closed system of truth functions published as [Post, 1941], a series of eighteen volumes on "Creative Logic" written 1938-1952, a series of seventeen volumes on "Theory of Finite Processes" (Volume IV is missing), written 1944-1951, and two volumes on "Definability," written 1952-1954. In addition, there are notebooks on "Calculus of Finite Processes" 1944, "Complete Equivalence of Normal Set and Recursive Function Development" 1942-1945 and "The Logic of Mathematics."

In his notebooks, Post frequently makes a meticulous record of clock times and of time spent working. An example quoted by Grattan-Guinness is of a manuscript on the Laplace transform. Post noted that he was writing the paper on 19 September 1923, starting at 10:08 AM, and ending at 12:28 PM, with three running times noted within the text [Grattan-Guinness, 1990, p. 79]. Post's original attack of mania occurred when he was enormously excited about his new ideas in logic in 1921, and subsequently he attempted to avoid periods of great mental excitement both by severely restricting the time he spent on research, and by alternating between one research project and another. Martin Davis has described the onset of one of his manic episodes in 1947:

> With another student, I had begun an "honors" reading course on mathematical logic under Post's tutelage. We had just reached the "deduction theorem" in propositional calculus, when Post met us full of excitement about his new work. We did not meet again with Post that semester; we had witnessed the beginning of one of his manic episodes. Apparently it was precisely excitement that had the potential to push him over the edge [Davis, 1994, p. xxiv].

It is impossible not to feel deep admiration for Post's remarkable achievements in mathematical logic, accomplished while labouring under such a severe handicap.

ACKNOWLEDGEMENTS

My greatest debt is to Martin Davis, whose editorial work and writings on Post were indispensable in preparing this chapter. The biographical essay [Gleiser, 1980] is a very informative source on Post's life. The description of Post's papers was obtained from the web site of the American Philosophical Society, which is

also a reliable source of biographical information. The online MacTutor History of Mathematics archive at the University of St Andrews, Scotland was also helpful. The Aspray interview with Kleene was obtained through the online version of the transcripts of the Princeton mathematical oral history project, obtainable at the web site of the Seeley G. Mudd Manuscript Library.

I thank Martin Davis and Solomon Feferman for helpful comments on a draft of this article, and would also like to thank Earle Spamer of the American Philosophical Society for informative correspondence.

BIBLIOGRAPHY

[Aspray, 1984] William Aspray. Interview with Stephen C. Kleene and J. Barkley Rosser. Transcript of recorded interview 26 April 1984 included in the oral history project *The Princeton Mathematics Community in the 1930s*, Seeley G. Mudd Manuscript Library, Princeton University (available online at the URL `http://www.princeton.edu/mudd/math`), 1984.

[Bar-Hillel *et al.*, 1961] Y. Bar-Hillel, M. Perles, and E. Shamir. On formal properties of simple phrase structure grammars. *Zeitschrift für Phonetik, Sprachwissenschaft, und Kommunikationsforschung*, 14:143–172, 1961. Reprinted in [Luce *et al.*, 1965, pp. 75–104].

[Berlekamp *et al.*, 1982] Elwyn R. Berlekamp, John H. Conway, and Richard K. Guy. *Winning Ways for your mathematical plays. Volume 2: Games in Particular*. Academic Press, London, New York, Paris, San Diego, San Francisco, São Paulo, Sydney, Tokyo, Toronto, 1982.

[Bernays, 1926] Paul Bernays. Axiomatische Untersuchung des Aussagen-Kalkuls der "Principia Mathematica". *Mathematische Zeitschrift*, 25:305–320, 1926.

[Böhler *et al.*, 2003] Elmar Böhler, Nadia Creignou, Steffen Reith, and Heribert Vollmer. Playing with Boolean Blocks, Part I: Post's lattice with applications to complexity theory. *ACM SIGACT News*, 34:38–52, 2003.

[Boone, 1957a] William Boone. Certain simple unsolvable problems of group theory V. *Indagationes Mathematicae*, 19:22–27, 1957.

[Boone, 1957b] William Boone. Certain simple unsolvable problems of group theory VI. *Indagationes Mathematicae*, 19:227–232, 1957.

[Brouwer, 1924] L.E.J. Brouwer. Beweis, dass jede volle Funktion gleichmässig stetig ist. *Koninklijke Akademie van wetenschappen te Amsterdam, Proceedings of the section of sciences*, 27:189–193, 1924.

[Chomsky, 1956] Noam Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, IT-2:113–124, 1956. Reprinted in [Luce *et al.*, 1965, pp. 105–124].

[Church, 1936] Alonzo Church. An unsolvable problem of elementary number theory. *The American Journal of Mathematics*, 58:345–363, 1936.

[Crossley, 1975] J.N. Crossley. Reminiscences of logicians. In J.N. Crossley, editor, *Algebra and Logic: Papers from the 1974 Summer Research Institute of the Australian Mathematical Society, Monash University, Australia*, pages 1–62, Berlin, Heidelberg, New York, 1975. Springer-Verlag. Lecture Notes in Mathematics 450.

[Davis, 1950] Martin Davis. *On the theory of recursive unsolvability*. PhD thesis, Princeton University, 1950.

[Davis, 1958] Martin Davis. *Computability and Unsolvability*. McGraw-Hill Book Company, New York, Toronto, London, 1958.

[Davis, 1965] Martin Davis, editor. *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. Raven Press, Hewlett, New York, 1965.

[Davis, 1982] Martin Davis. Why Gödel didn't have Church's thesis. *Information and Control*, 54:3–24, 1982.

[Davis, 1994] Martin Davis. Emil L. Post: His life and work. In *Solvability, Provability, Definability: The Collected Works of Emil L. Post*, pages xi–xxviii. Birkhäuser, Boston, Basel, Berlin, 1994.

[Epstein, 1960] George Epstein. The lattice theory of Post algebras. *Transactions of the American Mathematical Society*, 95:300–317, 1960.

[Feferman, 1995] Solomon Feferman. Turing in the land of O(z). In Rolf Herken, editor, *The Universal Turing Machine: A Half-Century Survey*, pages 103–134. Springer-Verlag, Wien, New York, 1995. Second edition.

[Friedberg, 1957] R.M. Friedberg. Two recursively enumerable sets of incomparable degrees of unsolvability. *Proceedings of the National Academy of Sciences of the U.S.A.*, 43:236–238, 1957.

[Ginsburg, 1966] Seymour Ginsburg. *The Mathematical Theory of Context-free Languages*. McGraw-Hill, New York, San Francisco, St. Louis, Toronto, London, Sydney, 1966.

[Gleiser, 1980] Molly Gleiser. The emotional logician. *Datamation*, 26, Number 3:221–223, 1980.

[Gödel, 1931] Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.

[Gödel, 1938] Kurt Gödel. The consistency of the axiom of choice and of the generalized continuum hypothesis. *Proceedings of the National Academy of Sciences, U.S.A.*, 24:556–557, 1938. In [Gödel, 1990, pp. 26–27].

[Gödel, 1947] Kurt Gödel. What is Cantor's continuum problem? *American Mathematical Monthly*, 54:515–525, 1947. In [Gödel, 1990, pp. 176–18727].

[Gödel, 1965] Kurt Gödel. Remarks before the Princeton bicentennial conference on problems in mathematics. In Martin Davis, editor, *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, pages 84–88. Raven Press, Hewlett, New York, 1965. In [Gödel, 1990, pp. 150–153].

[Gödel, 1990] Kurt Gödel. *Collected Works, Volume II: Publications 1938-1974*. Oxford University Press, New York, Oxford, 1990. Edited by Solomon Feferman, John W. Dawson Jr., Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay and Jean van Heijenoort.

[Gödel, 1995] Kurt Gödel. *Collected Works, Volume III: Unpublished Essays and Lectures*. Oxford University Press, New York, Oxford, 1995. Edited by Solomon Feferman, John W. Dawson Jr., Warren Goldfarb, Charles Parsons and Robert M. Solovay.

[Gödel, 2003] Kurt Gödel. *Collected Works, Volume V: Correspondence H-Z*. Oxford University Press, New York, Oxford, 2003. Edited by Solomon Feferman, John W. Dawson Jr., Warren Goldfarb, Charles Parsons and Wilfried Sieg.

[Goldfarb, 1981] Warren D. Goldfarb. The undecidability of the second-order unification problem. *Theoretical Computer Science*, 13:225–230, 1981.

[Grattan-Guinness, 1990] Ivor Grattan-Guinness. The manuscripts of Emil L. Post. *History and Philosophy of Logic*, 11:77–83, 1990.

[Herbrand, 1930] Jacques Herbrand. Les bases de la logique Hilbertienne. *Revue de Métaphysique et de Morale*, 37:243–255, 1930. Reprinted in [Herbrand, 1968, pp. 155–166].

[Herbrand, 1968] Jacques Herbrand. *Écrits Logiques*. Presses Universitaires de France, Paris, 1968. Bibliothèque de Philosophie Contemporaine.

[Hilbert and Ackermann, 1928] David Hilbert and Wilhelm Ackermann. *Grundzüge der theoretischen Logik*. Springer Verlag, Berlin, 1928.

[Hilbert, 1902] David Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 8:437–479, 1902. Reprinted [Yandell, 2002, pp. 389-424].

[Hopcroft and Ullman, 1979] John. E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, Massachussetts, 1979.

[Jevons, 1864] W. Stanley Jevons. *Pure Logic, or the logic of quality apart from quantity: with remarks on Boole's system and on the relation of logic and mathematics*. E. Stanford, London, 1864. Reprinted in [Jevons, 1890, pp. 3–77].

[Jevons, 1890] W. Stanley Jevons. *Pure Logic and other minor works*. MacMillan and Co., London and New York, 1890. Edited by Robert Adamson and Harriet A. Jevons. Reprinted by Thoemmes Press, Bristol 1991.

[Jockusch and Soare, 1973] Carl Jockusch and Robert Soare. Post's problem and his hypersimple set. *Journal of Symbolic Logic*, 38:446–452, 1973.

[Keyser, 1904] Cassius J. Keyser. The axiom of infinity: A new presupposition of thought. *Hibbert Journal*, 2:532–552, 1904.

[Kleene and Post, 1954] S.C. Kleene and Emil L. Post. The upper semi-lattice of degrees of unsolvability. *Annals of Mathematics*, 59:379–407, 1954.

[Kleene, 1936] S.C. Kleene. General recursive functions of natural numbers. *Mathematische Annalen*, 112:727–742, 1936.

[Kleene, 1943] S.C. Kleene. Recursive predicates and quantifiers. *Transactions of the American Mathematical Society*, 53:41–73, 1943.

[Kleene, 1952a] S.C. Kleene. *Introduction to Metamathematics*. Van Nostrand, Princeton, Toronto, London, 1952.

[Kleene, 1952b] S.C. Kleene. Recursive functions and intuitionistic mathematics. In *Proceedings of the International Congress of Mathematicians, Cambridge, Massachusetts, USA, 1950*, volume 2, pages 679–685, 1952.

[Lewis, 1918] C.I. Lewis. *A survey of symbolic logic*. University of California Press, Berkeley, 1918. Semicentennial Publications of the University of California.

[Linial and Post, 1949] Samuel Linial and Emil L. Post. Recursive unsolvability of the deducibility, Tarski's completeness, and independence of axioms problems of propositional calculus. *Bulletin of the American Mathematical Society*, 55:50, 1949. (Abstract), [Post, 1994, p. 550].

[Lorentz, 2001] R.J. Lorentz. Creating difficult instances of the Post correspondence problem. In Tony Marsland and Ian Frank, editors, *Second International Conference on Computers and Games (CG 2001), Hamamatsu, Japan*, pages 214–228. Springer Verlag, 2001. Lecture Notes in Computer Science 2063.

[Lucas, 1961] J.R. Lucas. Minds, machines and Gödel. *Philosophy*, 36:112–127, 1961.

[Luce *et al.*, 1965] R. Duncan Luce, Robert R. Bush, and Eugene Galanter, editors. *Readings in Mathematical Psychology Volume II*. John Wiley and Sons, New York, London, Sydney, 1965.

[Łukasiewicz, 1920] J. Łukasiewicz. O logice trójwartościowej (on three-valued logic). *Ruch filozoficzny*, 5:169–171, 1920. English translation in [Łukasiewicz, 1970].

[Łukasiewicz, 1970] J. Łukasiewicz. *Selected Works*. North-Holland, Amsterdam, Amsterdam, London, 1970. Edited by L. Borkowski; translated from the Polish by O. Wojtasiewicz.

[Lyndon, 1951] R.C. Lyndon. Identities in two-valued calculi. *Transactions of the American Mathematical Society*, 71:457–465, 1951.

[Markov, 1947] A.A. Markov. Impossibility of certain algorithms in the theory of associative systems (Russian). *Doklady Akademii Nauk SSSR (Reports of the Academy of Sciences of the USSR)*, 55:587–590, 1947.

[Matiyasevich, 1970] Yuri Matiyasevich. Enumerable sets are diophantine. *Soviet Mathematics Doklady*, 11:354–358, 1970.

[Matiyasevich, 1993] Yuri Matiyasevich. *Hilbert's Tenth Problem*. MIT Press, Cambridge, Massachusetts and London, 1993.

[Minsky, 1961] Marvin Minsky. Recursive unsolvability of Post's problem of "tag" and other topics in theory of Turing machines. *Annals of Mathematics*, 74:437–455, 1961.

[Minsky, 1967] Marvin L. Minsky. *Computation, finite and infinite machines*. Prentice-Hall, Englewood Cliffs, N.J., 1967.

[Muchnik, 1956] A.A. Muchnik. On the unsolvability of the problem of reducibility in the theory of algorithms (Russian). *Doklady Akademii Nauk SSSR (Reports of the Academy of Sciences of the USSR)*, 108:194–197, 1956.

[Murawski, 1998] Roman Murawski. E.L. Post and the development of mathematical logic and recursion theory. *Studies in Logic, Grammar and Rhetoric*, 2:17–30, 1998.

[Myhill and Scott, 1971] John Myhill and Dana Scott. Ordinal definability. *Proceedings of Symposia in Pure Mathematics*, 13, Part 1:271–278, 1971.

[Novikov, 1955] P.S. Novikov. On the algorithmic unsolvability of the word problem in group theory (Russian). *Trudy Ordena Lenina Matematicheskogo Instituta imeni V.A. Steklova. Akademiya Nauk SSSR (Proceedings of the Mathematical Steklov Institute of the Academy of Sciences SSSR)*, 44:1–143, 1955.

[Odifreddi, 1989] Piergiorgio Odifreddi. *Classical Recursion Theory: The Theory of Functions and Sets of Natural Numbers*. North-Holland, Amsterdam, New York, Oxford, Tokyo, 1989. Studies in Logic and the Foundations of Mathematics, Volume 125.

[Odifreddi, 1999] Piergiorgio Odifreddi. *Classical Recursion Theory Volume II*. Elsevier North Holland, Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore, Tokyo, 1999. Studies in Logic and the Foundations of Mathematics, Volume 143.

[Pelletier and Martin, 1990] Francis Jeffry Pelletier and Norman M. Martin. Post's functional completeness theorem. *Notre Dame Journal of Formal Logic*, 31:462–475, 1990.

[Penrose, 1989] Roger Penrose. *The Emperor's New Mind*. Oxford University Press, New York, Oxford, 1989.
[Pippenger, 1997] Nicholas Pippenger. *Theories of Computability*. Cambridge University Press, Cambridge, New York, 1997.
[Post, 1921a] Emil L. Post. Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, 43:163–185, 1921. [Post, 1994, pp. 21-43].
[Post, 1921b] Emil L. Post. On a simple class of deductive systems. *Bulletin of the American Mathematical Society*, 27:396–397, 1921. (Abstract), [Post, 1994, p. 545].
[Post, 1930] Emil L. Post. Generalized differentiation. *Transactions of the American Mathematical Society*, 32:723–781, 1930. [Post, 1994, pp. 44-102].
[Post, 1936] Emil L. Post. Finite combinatory processes – Formulation I. *Journal of Symbolic Logic*, 1:103–105, 1936. [Davis, 1965, pp. 289-291] [Post, 1994, pp. 103-105].
[Post, 1941] Emil L. Post. *The Two-Valued Iterative Systems of Mathematical Logic*. Princeton University Press, Princeton, New Jersey, 1941. Annals of Mathematics Studies, No. 5 [Post, 1994, pp. 249-374].
[Post, 1943] Emil L. Post. Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics*, 65:197–215, 1943. [Post, 1994, pp. 442-460].
[Post, 1944] Emil L. Post. Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society*, 50:284–316, 1944. [Davis, 1965, pp. 305-337] [Post, 1994, pp. 461-494].
[Post, 1946] Emil L. Post. A variant of a recursively unsolvable problem. *Bulletin of the American Mathematical Society*, 52:264–268, 1946. [Post, 1994, pp. 495-500].
[Post, 1948] Emil L. Post. Degrees of recursive unsolvability. Preliminary report. *Bulletin of the American Mathematical Society*, 54:641–642, 1948. (Abstract), [Post, 1994, pp. 549-550].
[Post, 1953a] Emil L. Post. A necessary condition for definability for transfinite von Neumann-Gödel set theory sets, with an application to the problem of the existence of a definable well-ordering of the continuum. Preliminary report. *Bulletin of the American Mathematical Society*, 59:246, 1953. (Abstract), [Post, 1994, pp. 551-552].
[Post, 1953b] Emil L. Post. Solvability, definability, provability: history of an error. *Bulletin of the American Mathematical Society*, 59:245–246, 1953. (Abstract), [Post, 1994, p. 551].
[Post, 1965] Emil L. Post. Absolutely unsolvable problems and relatively undecidable propositions – account of an anticipation. In Martin Davis, editor, *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, pages 340–433. Raven Press, Hewlett, New York, 1965. [Post, 1994, pp. 375-441].
[Post, 1990] Emil L. Post. The modern paradoxes. *History and Philosophy of Logic*, 11:85–91, 1990. Edited by I. Grattan-Guinness.
[Post, 1994] Emil L. Post. *Solvability, Provability, Definability: The Collected Works of Emil L. Post*. Birkhäuser, Boston, Basel, Berlin, 1994. Edited by Martin Davis.
[Reid, 1996] Constance Reid. *Julia, a Life in Mathematics*. Mathematical Association of America, Washington, DC, 1996. MAA Spectrum Series.
[Robinson, 1965] J.A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the Association for Computing Machinery*, 12:23–41, 1965.
[Rogers, 1967] Hartley Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill Book Company, New York, St. Louis, San Francisco, Toronto, London, Sydney, 1967. MIT Press Paperback Edition, 1987.
[Rosenbloom, 1942] Paul C. Rosenbloom. Post algebras. I. Postulates and general theory. *American Journal of Mathematics*, 64:167–188, 1942.
[Rosenbloom, 1950] Paul C. Rosenbloom. *The Elements of Mathematical Logic*. Dover Publications, Inc., New York, 1950.
[Rosser, 1936] J.B. Rosser. Extensions of some theorems of Gödel and Church. *Journal of Symbolic Logic*, 1:87–91, 1936.
[Russell, 1904] Bertrand Russell. The axiom of infinity. *The Hibbert Journal*, 2:809–812, 1904.
[Shoenfield, 1971] Joseph R. Shoenfield. *Degrees of Unsolvability*. North-Holland Publishing Company: American Elsevier Publishing Company, Amsterdam, London, New York, 1971.
[Sieg, 1997] Wilfried Sieg. Step by recursive step: Church's analysis of effective calculability. *Bulletin of Symbolic Logic*, 3:154–180, 1997.
[Soare, 1987] Robert I. Soare. *Recursively Enumerable Sets and Degrees: A Study of Computable Functions and Computably Generated Sets*. Springer-Verlag, Berlin, Heidelberg, New York, 1987.

[Takeuti, 1961] Gaisi Takeuti. Remarks on Cantor's absolute. *Journal of the Mathematical Society of Japan*, 13:197–206, 1961.

[Thue, 1914] Axel Thue. Problem über Veränderungen von Zeichenreihen nach gegebenen Regeln. *Skrifter utgit av Videnskapsselskapet i Kristiania, I. Matematisk-naturvidenskabelig klasse, No. 10*, 1914.

[Traczyk, 1963] T. Traczyk. Axioms and some properties of Post algebras. *Colloquium Mathematicae*, 10:193–209, 1963.

[Traczyk, 1964] T. Traczyk. An equational definition of a class of Post algebras. *Bulletin de l'Académie Polonaise des Sciences. Séries des Sciences Mathématiques, Astronomiques et Physiques*, 12:147–150, 1964.

[Tseitin, 1958] G.S. Tseitin. An associative calculus with an unsolvable problem of equivalence (Russian). *Trudy Ordena Lenina Matematicheskogo Instituta imeni V.A. Steklova. Akademiya Nauk SSSR (Proceedings of the Mathematical Steklov Institute of the Academy of Sciences SSSR)*, 52:173–189, 1958. English translation *American Mathematical Society Translation Series 2*, Volume 94 (1970), pp. 73-92.

[Turing, 1936] Alan Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, Series 2*, 42:230–265, 1936.

[Turing, 1939] Alan Turing. Systems of logic based on ordinals. *Proceedings of the London Mathematical Society, Series 2*, 45:161–228, 1939. [Davis, 1965, pp. 155-222].

[Uspensky, 1983] V.A. Uspensky. *Post's Machine*. MIR Publishers, Moscow, 1983. Translated from the 1979 Russian edition by R. Alavina.

[Vaughan, 1941] H.E. Vaughan. Review of [Post, 1941]. *Journal of Symbolic Logic*, 6:114–115, 1941.

[Vollmer, 1999] Heribert Vollmer. *Introduction to Circuit Complexity: A Uniform Approach*. Springer-Verlag, Berlin, Heidelberg, New York, 1999. Texts in Theoretical Computer Science: An EATCS Series.

[von Neumann, 1927] John von Neumann. Zur Hilbertschen Beweistheorie. *Mathematische Zeitschrift*, 26:1–46, 1927.

[Whitehead and Russell, 1910-13] Alfred North Whitehead and Bertrand Russell. *Principia Mathematica*. Cambridge University Press, Cambridge, 1910-13. Volume 1 1910, Volume 2 1913, Volume 3 1913.

[Wittgenstein, 1921] Ludwig Wittgenstein. Logisch-philosophische Abhandlung. *Annalen der Naturphilosophie*, 14:185–262, 1921. With an introduction (in German) by Bertrand Russell.

[Wittgenstein, 1922] Ludwig Wittgenstein. *Tractatus Logico-Philosophicus*. Routledge and Kegan Paul, London and New York, 1922. English translation of [Wittgenstein, 1921], with an introduction by Bertrand Russell.

[Wolfram, 2002] Stephen Wolfram. *A New Kind of Science*. Wolfram Media Inc., Champaign, IL, 2002.

[Yandell, 2002] Benjamin H. Yandell. *The Honors Class: Hilbert's Problems and Their Solvers*. A K Peters, Natick, Massachusetts, 2002.

[Zhao, 2002] Ling Zhao. Solving and creating difficult instances of Post's correspondence problem. Master's thesis, University of Alberta, 2002.

GENTZEN'S LOGIC

Jan von Plato

1 GENTZEN'S LIFE AND WORK IN BRIEF

Gerhard Gentzen was born on 24 November 1909 in Greifswald, a northern German city by the Baltic Sea. He began his studies in mathematics and physics there in 1928, then moved to Göttingen in 1929. His doctoral thesis *Untersuchungen über das logische Schliessen* (Investigations into logical deduction), under the supervision of Paul Bernays, was accepted in June 1933 and published in two parts in 1934–35. It contained Gentzen's main contribution to the study of pure logic, the systems of intuitionistic natural deduction and classical sequent calculus, and marked the birth of structural proof theory. National socialist racial laws forced Bernays to leave his professorship at Göttingen in the spring of 1933, even before the examination of the doctoral thesis, but the two were able to keep up correspondence until the beginning of the Second World War. During this period, Gentzen worked in Göttingen and behaved overtly like any person in accord with the Nazi regime – not that he would have convinced anyone involved with the latter. The continued private correspondence with an expelled Jewish professor was certainly enough to reveal the nature of his attitude.

The proof-theoretical study of pure logic was a preliminary to Gentzen's greater aims, namely consistency proofs for arithmetic and analysis. The Gödel-Gentzen translations of classical Peano arithmetic into intuitionistic Heyting arithmetic had shown that, Gödel's second incompleteness theorem notwithstanding, the consistency of arithmetic is within the reach of constructive reasoning. Following this conviction, Gentzen found a proof of consistency late in 1934, for a system that used natural deduction. The published version of 1936 had a different proof based on the famous principle of transfinite induction up to the ordinal ϵ_0 by which consistency followed. A third proof of 1938 used transfinite induction but the logical system was a sequent calculus.

Gentzen's last work, about the "provable initial segments of transfinite induction," was finished in 1939. It gave a direct proof of the unprovability of ϵ_0-induction in an extension of Peano arithmetic with transfinite ordinals. The induction principle is expressible in arithmetic, and therefore Gödel's first incompleteness theorem follows directly, without the metamathematical coding of the provability predicate as in Gödel's original work. Secondly, yet another, fourth proof of consistency followed as an immediate consequence of the unprovability

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

of transfinite induction. These two results clarified rather fully the situation created by the incompleteness of arithmetic. At the same time, Gentzen's last paper marked the beginning of ordinal proof theory. Gentzen's efforts at proving the consistency of the theory of real numbers came to an end through his untimely and tragic death in 1945, at the age of thirty-five. Freed from military service early in 1942 because of nervous exhaustion, he had become a teacher at the German university of Prague in 1943. At the end of the war, the newly organized, Soviet-backed local police of the city arrested the German professors. Gentzen was unable to do the required physical work because of a hand injury, was therefore deprived of food, and died of hunger in prison on 4 August 1945.

The sources of the following presentation, besides Gentzen's published papers and manuscripts, are almost all documented in Eckart Menzler-Trott's pioneering biography of Gentzen, *Gentzens Problem* (2001), to appear in English translation as *Logic's Lost Genius* (2007). The notation is uniform and what one would write today in a logic text in English, but it is so close to the original that a collateral reading of passages from Gentzen's papers can be done without difficulty. We refer to passages in Gentzen's papers by the detailed numbering system he used, rather than page numbering.

The references section lists all of the eleven articles Gentzen is known to have written for publication. He wrote also a number of short reviews and abstracts that can be found described in Menzler-Trott's biography.

It was believed until the mid-1980s that all of Gentzen's unpublished papers had disappeared, but then it was instead found out that his sister was in the possession of manuscripts. These are written in shorthand and record work done between 1931 and 1944. The Gentzen papers are being laboriously transcribed for eventual publication by Christian Thiel of Erlangen University. He has generously let us study both the transcribed and the manuscript parts as well as other documents relating to Gentzen for work on this *Handbook* chapter; a brief description of the contents of the manuscripts is given in the last Section. Thiel has also a summary Gentzen prepared of his work on the consistency of analysis, dated February 1945. Lastly, we located in February 2005 an early handwritten manuscript version of Gentzen's doctoral thesis. It had been, so to speak, in front of everyone's eyes since 1969, because its second page is reproduced photographically on page *xiii* of Gentzen's *Collected Papers*. Remarkably, this manuscript contains a detailed proof of normalization for intuitionistic natural deduction, absent from the printed version.

Most of the presentation that follows is concentrated on historical aspects of Gentzen's achievement, including the later development of lines of research that he initiated. In keeping with the character (and title) of this *Handbook*, we enter at places into systematic foundational discussions, especially when we have some new points of view to present, such as on the correspondence between natural deduction and sequent calculus, or on the application of sequent calculus in mathematics.

My knowledge of Gentzen's life and work, beyond his published papers, is mostly based on the efforts of Eckart Menzler-Trott. He collected for over two decades,

with unfading enthusiasm, facts and documents about Gentzen, and posed questions in correspondence to persons many of whom are deceased long since. I dedicate this article to Eckart, with the wish that my fellow academic historians of logic take his work for what it is: that of an *amateur par excellence*.

2 GENTZEN'S FIRST PAPER

Paul Hertz, otherwise known for his contribution to the foundations of statistical mechanics, put up in the 1920s a general theory of "systems of sentences" (Hertz 1929). A *sentence* is an expression of the form $a_1, \ldots, a_n \rightarrow b$. It has several interpretations: the circumstances given by $a_1, \ldots, a_n$ bring about b, objects with the properties $a_1, \ldots, a_n$ have also the property b, etc. The logical interpretation is that from the propositions $a_1, \ldots, a_n$, proposition b follows. The arrow thus represents the notion of logical consequence.

Hertz systems have two forms of rules of inference, the first with zero premisses:

$$a_1, \ldots, a_n \rightarrow a_i \quad \text{for } 1 \leqslant i \leqslant n,$$

$$\frac{a_{1_1}, \ldots, a_{1_n} \rightarrow b_1 \ \ldots \ a_{m_1}, \ldots, a_{m_k} \rightarrow b_m \quad b_1, \ldots, b_m, c_1, \ldots, c_l \rightarrow c}{a_{1_1}, \ldots, a_{1_n}, \ldots, a_{m_1}, \ldots, a_{m_k}, c_1, \ldots, c_l \rightarrow c}$$

Table 1. The rules of Hertz systems.

Hertz called the former kind of rules "immediate inferences" and the latter "syllogisms." If a collection of sentences S is closed with respect to these rules, it forms a *sentence system*. A collection of sentences T is an *axiom system* for a sentence system S if each sentence in S is derivable from the ones in T.

One of the principal questions for Hertz was the existence of independent axiom systems, with independence defined in the obvious way, as underivability by the rules. Bernays had followed and clearly also sustained the work of Hertz, and he suggested that Gentzen work with the problem. The result was Gentzen's first paper, titled, in English translation, "On the existence of independent axiom systems for infinitary sentence systems." He showed by a counterexample that there is an infinite sentence system that permits no axiom system with independent axioms.

Gentzen noticed in the course of his work, conducted in the summer of 1931, that the second form of Hertz' rules can be replaced by one in which there is just one term b as a middle term of syllogism:

$$\frac{a_1, \ldots, a_m, \rightarrow b \quad b, c_1, \ldots, c_n \rightarrow c}{a_1, \ldots, a_m, c_1, \ldots, c_n \rightarrow c}$$

Table 2. Gentzen's cut rule in Hertz systems.

He called this form of rule *cut* (Schnitt), and the first form of Hertz' rules *thinning* (Verdünnung). The "sentences" of the calculus of Hertz became the "sequents" of Gentzen's later work.

The calculus of Hertz was one of the two main sources of Gentzen's sequent calculus. The other source was his development of the calculus of natural deduction. The rules of the latter, when adapted to the notation of sequents, led to sequent calculus proper as we shall see in Section 8. The background of sequent calculus in the work of Hertz and in Gentzen's first paper is discussed at length in Bernays (1965).

Gentzen's work on Hertz systems can be seen as an early contribution to logic programming. The sentences are the same as program clauses, and derivations consist of cuts with the terms in the clauses. Gentzen's work is described from this point of view in Peter Schroeder-Heister's paper (2002). The main results contain a normal form for derivations that amounts, in modern terms, "to the completeness of propositional SLD-resolution in logic programming" (ibid., p. 246).

Gentzen soon abandoned the topic of his first paper. He writes on 13 December 1932 to his old professor in Greifswald, Hellmuth Kneser (from Mentzler-Trott 2001, p. 35):

> I have set as my specific task to find a proof of the consistency of logical deduction in arithmetic... The task becomes a purely mathematical problem through the formalization of logical deduction. The proof of consistency has been so far carried out only for special cases, for example, the arithmetic of the integers without the rule of complete induction. I would like to proceed further at this point and to clear at least arithmetic with complete induction. I am working on this since almost a year and hope to finish soon, and would then present this work as my dissertation (with Prof. Bernays).

3 INTERPRETATION OF PEANO ARITHMETIC IN HEYTING ARITHMETIC

When intuitionistic logic was in its infancy in the late 1920s, a lively debate arose on its proper formalization. It was conducted on the pages of the *Bulletin* of the Royal Belgian Academy. Valeri Glivenko contributed to this discussion in 1929 by a result by which classically provable negative formulas of propositional logic are already intuitionistically provable. The origins of the idea go back to Andrei Kolmogorov's paper of 1925; In it, Kolmogorov gave a double-negation interpretation of classical logic in intuitionistic logic. For each formula A, let A^* stand for the formula obtained by prefixing each subformula of A with a double negation. If A is provable classically, then A^* is provable intuitionistically.

Kolmogorov's paper was written in Russian and remained unknown to logicians outside Moscow, except for its indirect influence through Glivenko (1929). In 1932, Gödel and Gentzen worked out independently of each other an interpretation of classical Peano arithmetic in the intuitionistic arithmetic of Heyting. Gentzen's translation was as follows:

1. $a = b^* \quad \rightsquigarrow \quad a = b$,

2. $(\neg A)^* \quad \rightsquigarrow \quad \neg A^*$,
3. $(A\&B)^* \quad \rightsquigarrow \quad A^*\&B^*$,
4. $(A \supset B)^* \quad \rightsquigarrow \quad A^* \supset B^*$,
5. $(\forall x A)^* \quad \rightsquigarrow \quad \forall x A^*$,
6. $(A \vee B)^* \quad \rightsquigarrow \quad \neg(\neg A^* \& \neg B^*)$,
7. $(\exists x A)^* \quad \rightsquigarrow \quad \neg\forall x \neg A^*$.

Table 3. The Gödel-Gentzen translation.

Gentzen tells to add double-negations to equations if these contain variables. Gödel's translation removes also implications $A \supset B$, by translating them into $\neg(A^*\&\neg B^*)$. The crucial points of the translation are disjunction and existence. The former is translated into the intuitionistically weaker $\neg(\neg A^* \,\&\, \neg B^*)$, and the latter into $\neg\forall x \neg A^*$. Even though there is a difference in the translations Gödel and Gentzen defined, the translation theorem that they established was the same: A formula A is a theorem of classical Peano arithmetic if and only if A^* is a theorem of intuitionistic Heyting arithmetic.

Indirect inferences with the quantifiers were thought to be the "dubious" transfinite element in mathematics, to be saved by a finitary proof of consistency as required by Hilbert's program. By the translation, the consistency of classical arithmetic reduces to that of intuitionistic arithmetic, so that indirect inferences in arithmetic are legitimate. Put in another way: The abandonment of indirect proofs has no real effect, because formulas that are provable only indirectly can be substituted by classically equivalent intuitionistically provable ones.

As a special case of the Gödel-Gentzen translation, with all atomic formulas double-negated, a translation of classical predicate logic into intuitionistic predicate logic is obtained.

Heyting corresponded with Gentzen in the winter and spring of 1933. Only the letters of the latter have been preserved. It appears from these letters that Gentzen had sent his paper *Über das Verhältnis zwischen intuitionistischer und klassischer Arithmetik* (On the relation between intuitionistic and classical arithmetic) to Heyting, and also explained some results that appeared later in the doctoral thesis. Heyting then wrote to Gödel about Gentzen's interpretation of classical arithmetic in intuitionistic arithmetic, with Gödel answering on 16 May that he had found the translation and that "it should have been known in Göttingen since at least June 1932." As a reason for the latter, Gödel wrote that at that time he had "reported on it in the Menger colloquium, to wit, in the presence of O. Veblen who shortly afterwards went to Göttingen." (See the Gödel-Heyting correspondence in volume IV of Gödel's *Collected Works*.) We know from Menger's later recollections, based on minutes of the meetings, that Veblen got very enthusiastic about the lecture, and invited Gödel to visit the newly established Institute for Advanced Study in Princeton. However, there is no reason to think that Gentzen would have known of Gödel's discovery. It seems that at least three persons had the

idea, the first and most original one being Kolmogorov in 1925. Gentzen may have taken his inspiration from Glivenko's theorem and therefore, indirectly, from Kolmogorov, as Gödel did (see Troelstra's introduction to Gödel's paper in vol. I of Gödel's *Collected Works*, p. 282). Upon the publication of Gödel's paper, Gentzen withdrew his own paper, already at proof stage, from publication. Luckily these proofs were preserved by Bernays, and they appeared in an English translation in 1969, in Manfred Szabo's edition of Gentzen's *Collected Papers*, then in the original German in 1974. We shall see in Section 5 that Gentzen's paper was taken out directly from the early version of his doctoral thesis.

The importance of the Gödel-Gentzen translation was that it clearly showed the consistency of arithmetic to be within the reach of intuitionistic arguments. A letter Gentzen wrote to Heyting on 25 February 1933 summarizes the situation as follows:

> A consistency proof by finite means has ... not succeeded so far, so that this original aim of Hilbert has not been achieved. If, on the other hand, one admits the intuitionistic position as a secure basis in itself, i.e., as a consistent one, the consistency of classical arithmetic is secured by my result. If one wished to satisfy Hilbert's requirements, the task would still remain of showing intuitionistic arithmetic consistent. This, however, is not possible by even the formal apparatus of classical arithmetic, on the basis of Gödel's result in combination with my proof. Even so, I am inclined to believe that a consistency proof for intuitionistic arithmetic, from an even more evident position, is possible and also desirable.

The last-mentioned was the goal Gentzen had set to himself early in 1932.

4 NATURAL DEDUCTION

The first task in Gentzen's consistency program was to study pure logic proof-theoretically. Gentzen's previous experience in logic included the use of systems of axiomatic logic in the tradition of Frege and Russell. Hilbert and Ackermann's little book *Grundzüge der theoretischen Logik* of 1928 presented classical propositional and predicate logic axiomatically. Heyting's pioneering work of 1930 did the same for intuitionistic logic.

Of his motivations for abandoning the axiomatic tradition in logic, Gentzen says little. He just went on, as he writes in his doctoral thesis, to construct a system of logic that is "as close to actual reasoning as possible." The main feature of such reasoning is its hypothetical character: A mathematical argument starts with some given assumptions that are analyzed into parts. At some stage, a conclusion is synthesized from suitable parts. Whenever the assumptions are verified, also the conclusion is. The result of Gentzen's exercise was his calculus of *natural deduction*. We shall follow first the presentation of the published thesis, then report on the newly discovered manuscript version.

The preliminaries to Gentzen (1934–35) contain a general definition of *derivations*: An *inference figure* can be written in the form

$$\frac{A_1 \ \ldots \ A_n}{B}$$

where $n \geqslant 1$. A *derivation* "consists of formulas (at least one) that form among themselves inference figures as follows: Each formula is the conclusion of at most one figure, and each formula except for the *endformula* is the premiss of at least one figure, and the system of figures is acyclic." (1934–35, I.3.2). Acyclicity requires that there be no repetition of a formula in any premiss-conclusion sequence. Further explanations specify that the clauses concern formula occurrences in a derivation.

If the same premiss formula occurs in several inference figures, the combination of these two-dimensional figures need not have any simple representation in two dimensions. It should be clear that no two-dimensional graphical representation of derivations is intended by the definition, but an abstract combinatorial structure. The condition of acyclicity is a further indication of a combinatorial rather than geometric concept of derivations. In the former, but not in the latter, it is needed in order to rule out infinite premiss-conclusion sequences in derivations.

Gentzen's carefully laid definition contains as a special case formulas that are not conclusions of any figure, by the "at most" clause on conclusions. These are the *topformulas* of a derivation. A derivation is next defined to be *of tree form* if each formula is the premiss of *at most* one figure. Without this condition, derivations have the structure of directed acyclic graphs, in today's terminology.

Gentzen's analysis of the practice of mathematical proofs led him to classify steps of proof into two types: ones that give prescriptions of the sufficient grounds for concluding a proposition of a given form, the *introduction* rules of natural deduction, and secondly steps of proof that give necessary consequences of a proposition of a given form, the *elimination* rules. The former are as follows:

$$\frac{A \quad B}{A\&B}\,{\&I} \qquad \frac{A}{A \vee B}\,{\vee I_1} \qquad \frac{B}{A \vee B}\,{\vee I_2} \qquad \frac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \supset B}\,{\supset I} \qquad \frac{A(y/x)}{\forall x A}\,{\forall I} \qquad \frac{A(t/x)}{\exists x A}\,{\exists I}$$

Table 4. The introduction rules of natural deduction.

In rule $\supset I$, any number of copies of formula A, the case of none at all included, can be *discharged*. Gentzen indicates these formulas in the schematic rule by the notation $[A]$. In actual derivations, a numerical label is put next to the rule and on top of those assumptions discharged by the rule. Rule $\forall I$ has the variable restriction: The eigenvariable y must not occur free in any assumptions the result of substitution $A(y/x)$ depends on. In rule $\exists I$, t is an arbitrary term. Gentzen writes formulas with possible free occurrences of a variable x, and the result of a

substitution of x by a term t, as Ax and At, respectively. The notation for the universal quantifier $\forall$, now always used, is introduced in Gentzen's thesis.

Elimination rules are reverses of sorts of the introduction rules:

$$\frac{A\&B}{A}\,\&E_1 \qquad \frac{A\&B}{B}\,\&E_2 \qquad \frac{A \vee B \quad \begin{array}{c}[A]\\ \vdots\\ C\end{array} \quad \begin{array}{c}[B]\\ \vdots\\ C\end{array}}{C}\,\vee E \qquad \frac{A \supset B \quad A}{B}\,\supset E$$

$$\frac{\forall x A}{A(t/x)}\,\forall E \qquad \frac{\exists x A \quad \begin{array}{c}[A(y/x)]\\ \vdots\\ C\end{array}}{C}\,\exists E$$

Table 5. Gentzen's elimination rules for natural deduction.

The same variable restriction holds for rule $\exists E$ as for rule $\forall I$, i.e., the eigenvariable y of $\exists E$ must be arbitrary. Derivations can be started with any formulas

$$A$$

that act as *open assumptions*. A careful reading of Gentzen's thesis shows that finished derivations have no open assumptions, i.e., these are all discharged in the end by rule $\supset I$.

Various ways of treating negation are mentioned in Gentzen's thesis. It can be a primitive connective, or it can be defined through implication and a constant formula $\perp$ denoting *falsity*, as in $\neg A \equiv A \supset \perp$. There is no introduction rule for $\perp$, but there is instead an elimination rule

$$\frac{\perp}{C}\,\perp E$$

Table 6. The rule of falsity elimination.

The system of rules of introduction and elimination defines *intuitionistic natural deduction*, or system *NJ* in Gentzen's nomenclature.

Of the relation between introduction and elimination rules, Gentzen writes that the introduction rules of natural deduction can act as operational definitions of the connectives. Elimination rules are, then, as Gentzen says, just consequences of these definitions: "In the elimination of a sign one can use the formula, the outermost sign of which is in question, only 'in its meaning as based on the introduction of this sign'." (1934–35, II.5.13). Gentzen's example of this situation is as follows: If $A \supset B$ is introduced, followed by an implication elimination concluding B from $A \supset B$ and A, it can be obtained by "concluding B immediately from A that has been proved, for $A \supset B$ documents the existence of a derivation of B from A" (1934–35, II.5.13). We note that the argument assumes that derivations of $A \supset B$ can be normalized. The same explanation of $A \supset B$ is given in (1936, 11.1).

Gentzen's observation is that steps of introduction followed by the corresponding eliminations are useless detours (*Umwege*) in derivations. His example is written formally as:

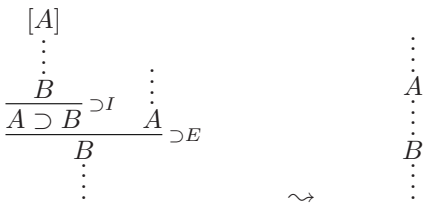


Table 7. A detour convertibility on implication.

In modern terms, an *I*-rule followed by the corresponding *E*-rule forms a *detour convertibility*. A derivation without such convertibilities is in *normal form*.

Gentzen writes that he was able to establish the existence of a normal form, for any given derivation of a conclusion C from open assumptions $A_1, \ldots, A_n$, only for the system of intuitionistic natural deduction, but not for the classical one. He therefore invented another logical calculus, that of sequents, and proved the analogue of the convertibility to normal form for the latter. The example of detour conversion on implication seems to be the only trace that is left in his paper of a proof of convertibility into normal form. Detour conversions for $\exists$ and for $\vee$ can be found in (1936, end of 10.3). Detour convertibilities are further described as "peaks of complexity" in (1938, 3.1).

The main property of derivations in normal form is the *subformula property*: All formulas in a normal derivation are subformulas of the open assumptions or of the conclusion. If the former are discharged by $\supset I$, all formulas are subformulas of the conclusion. For the reason mentioned, Gentzen goes on in (1934–35) to work out the central results of the analysis of logical proofs in terms of sequent calculi, for both intuitionistic and classical logic. An application of the intuitionistic sequent calculus was the decidability of intuitionistic propositional logic. The proof gives a bound to the size of possible derivation trees by showing that, in any derivation, the number of times a formula needs to occur on either side of a sequent is at most three. Another application was to the *disjunction property* of intuitionistic logic: If $A \vee B$ is derivable, a derivation of A or B can be found. In particular, there cannot be any proof of the law of excluded middle $A \vee \neg A$. A corresponding proof of the disjunction property directly in Gentzen's system of natural deduction would not be as straightforward as the proof in sequent calculus.

5 A NEWLY DISCOVERED PROOF OF NORMALIZATION BY GENTZEN

The Paul Bernays collection at ETH Zurich contains as one item a handwritten version of the *Untersuchungen über das logische Schliessen*, numbering 41 pages and with the, rather misleading, catalogue indication that the 12 first pages have been published as Gentzen's thesis in the *Mathematische Zeitschrift*. We shall refer

to this texs as *Ms. ULS*. It consists of various strata, also differentiated by the hand applied. The text begins with a one-page "setting out of the field of investigation." Next section I introduces the calculus *N1J* with the 1 dropped out in a later section and the printed version. In this part the handwriting is in the "deutsche Schrift," quite hard for a modern reader as one may see from the reproduced text in Szabo's edition. Now follows a part, beginning with page 13, in which the handwriting is more relaxed, but still in a style Gentzen would use when writing for others to read. The part is indicated as "Dissertation, III. Abschnitt. G. Gentzen." Next to these indications there is written, in a rather clumsy hand Gentzen would use when making notes to himself, "IV. Blatt." Gentzen used folded "Kanzleiformat" paper, high as a legal size and each sheet giving four pages. Gentzen gives a very detailed proof of normalization, written out as carefully as the materials in his published thesis. The proof is typical Gentzen: elaborate verbal arguments and a necessary minimum of formal notation. There then follows a brief section in which we see, so to say, sequent calculus being born in front of our eyes. Another short section introduces the multisuccedent sequent calculus. Between these pages there is a page of corrections numbered 33, but these corrections are exactly the ones found in Gentzen's paper (1933) discussed in Section 3 above. As a last item in the text comes a proof of the equivalence of Gentzen's calculus and an axiomatic calculus. This part is indicated as section II in the end of section I and is written in the same meticulous "deutsche" hand, numbered separately from page 1 on, with a casual added numbering from page 37 on.

Section I of *Ms. ULS* ends with a summary of results obtained so far, seemingly meant also to list the planned contents of the thesis: First, the equivalence of section II, with an auxiliary calculus bearing the nomenclature *N2*, and secondly the interpretation of classical arithmetic in intuitionistic arithmetic as section III. At some stage in 1932, Gentzen noticed that most of his rules of natural deduction have the property that the conclusion is a subformula of a premiss, or the other way around. There is thus another planned section IV with the following

> Conjecture of a theorem: If a logical proposition is provable in the calculus *N1J*, there is a proof for it in which only subformulas of it occur, possibly with different variables.

The text states further that the proof of this theorem is not finished, except for a calculus with rules for conjunction and the universal quantifier and a rule corresponding to negation introduction. Some time later Gentzen cancelled out the "conjecture" into "theorem" and declared the proof as finished. A last planned section V of *Ms. ULS* announces a proof of the consistency of arithmetic, one that "uses the concept of an infinite sequence of natural numbers and further, at one point, the law of excluded middle."

The *Ms. ULS* must have contained an original text of a section III on the interpretation of classical arithmetic in intuitionistic arithmetic. It was taken out to become, it seems *verbatim*, the paper Gentzen (1933), with just page 33 with corrections to this part left in between the preserved manuscript. Marginal remarks

indicate that the normalization proof was intended to replace the old section III of the thesis.

From the manuscript *Ms. ULS* and related materials, three successive plans for a doctoral thesis can thus be distinguished:

1. The original plan, which has a certain continuity of design to it, namely to: I introduce the system of intuitionistic natural deduction, II show that it is equivalent to an axiomatic intuitionistic calculus, III show that axiomatic calculi for classical logic and arithmetic can be interpreted in axiomatic calculi for intuitionistic logic and arithmetic, IV show that derivations in intuitionistic natural deduction are normalizing and have the subformula property, V prove that intuitionistic arithmetic is consistent.

2. In a second phase, around January 1933, the text of III became a separate paper and V was dropped out, with the likely motive that the planned approach through an extension of the method of section IV, i.e., normalization and the subformula property, from pure logic to derivations in arithmetic, did not work. Gentzen had sent the separate paper to Heyting who answered on 18 February 1933. Gentzen, in turn, wrote back on 25 February that he "hopes to be able to make investigations on the consistency of intuitionistic arithmetic next year." Thus, the hopes of the previous December, of "finishing soon" a consistency proof, as in the letter to Kneser, had faded in a little over a month.

3. In a last phase, the faint beginnings of intuitionistic and classical sequent calculi, added who knows when to the *Ms. ULS*, became the real topic of the doctoral thesis. The originally planned section IV on normalization, renumbered III in the second phase, was dropped out (and remained lost until the writing of this chapter in 2005), and section II on the equivalence with axiomatic calculi was placed last as in the printed thesis.

We shall now present the main novelties of *Ms. ULS*, if they can be so described, as compared to the printed thesis. An example of a formula, seen in the photographic reproduction of the top part of page 2 of section I in Szabo's edition, and different from the one in the published thesis, serves to introduce the idea of natural deduction. The remarkable feature of the treatment of the example is the adherence to the "linear" succession of steps of proof. The formula to be derived is $(A \vee B)\&(A \vee C) \rightarrow A \vee (B\&C)$. (Gentzen uses X, Y, and Z as well as dots for avoiding parentheses, as well as $\rightarrow$ for implication.) First one assumes $(A \vee B)\&(A \vee C)$, then $A \vee B$ is concluded. Next Gentzen considers simultaneously, on one line, A and B separately as a "division into cases." Both cases lead to the desired conclusion $A \vee (B\&C)$. The derivation does not quite succeed in displaying a linear arrangement of formulas in a derivation. The next example, the same as in the published version, concludes in a linear fashion from an existential formula an instance with an arbitrary term, instead of using the standard existence elimination pattern. This pattern is instead given on page 5 of *Ms. ULS*

that gives the rules in the form that has become standard. All three of Gentzen's examples are given formal derivations that follow strictly the rule system. In the presentation of this system, the following rules have been crossed out and replaced by the rule of falsity elimination: A "tautology" rule that concludes A from the premiss A, a rule R that concludes $\neg A$ if the assumption A led to contradictory conclusions, and a rule that concludes an arbitrary formula C from the premisses A and $\neg A$. The corrected system is exactly the one that is now called intuitionistic natural deduction. The notation (x) is used for the universal quantifier, similarly to the article Gentzen (1933) that was submitted in mid-March 1933, whereas the published thesis used $\forall$. A possible motive for the tautology rule, described as "trivial" by Gentzen, could be that the derivation of the formula $A \longrightarrow A$ is immediate with it. Otherwise one would usually say that the derivation of the premiss of rule $\rightarrow I$ has a limiting case in which the derivation is simply an assumption.

Page 9 of section I of *Ms.ULS* gives the standard detour convertibilities of natural deduction. For example, the implication conversion scheme is written as follows:

$$\dfrac{\alpha\quad\dfrac{\begin{matrix}[A]\\ \gamma\\ B\end{matrix}}{A\to B}\to I \qquad A}{\begin{matrix}B\\ \varepsilon\end{matrix}}\to E \qquad \text{becomes} \qquad \begin{matrix}\left[\begin{matrix}\alpha\\ A\end{matrix}\right]\\ \gamma\\ B\\ \varepsilon\end{matrix}$$

Table 8. A detour conversion scheme in Gentzen's manuscript.

The square bracket notation for any number of discharged assumptions A in rule $\rightarrow I$ is applied in the converted derivation also to indicate as many copies of the derivation of the minor premiss.

Gentzen's proof of normalization for intuitionistic natural deduction has been worked out directly for the whole calculus *N1J*. This involves a remarkable difficulty, as compared to a proof of normalization for a calculus without rules for disjunction and existence. Namely, an introduction-elimination pair can be separated by a succession of rules $\vee E$, $\exists E$. If, say, $C\&D$ is a minor premiss of $\vee E$ introduced in a last step, the conclusion of $\vee E$ is $C\&D$ and it can be the premiss of rule $\&E$. Rule $\vee E$ between $\&I$ and $\&E$ prevents conversion, as in

$$\dfrac{\dfrac{A\vee B \qquad \begin{matrix}[A]\\ \vdots\\ C\&D\end{matrix} \qquad \dfrac{\begin{matrix}[B]\\ \vdots\\ C\end{matrix}\quad\dfrac{\begin{matrix}[B]\\ \vdots\\ D\end{matrix}}{C\&D}\&I}{C\&D}\vee E}{C}\&E}{}$$

What is called a *permutation conversion* in Prawitz (1965) brings out such hidden detour convertibilities.

The margin to the beginning of the section with the normalization proof contains a list of changes: The calculus is to be called *NJ* instead of *N1J*, and the

terminology is to be derivations and formulas or expressions instead of proofs and propositions. Furthermore, the horseshoe is now used for implication, which definitely shows that the section is a later addition with respect to the other preserved parts of the *Ms. ULS*. With this change, the notation follows that of Heyting (1930).

Gentzen begins his proof by defining the notions of a *principal formula* of a rule, the one with the corresponding connective or quantifier. The subformulas of the principal formula that appear in a rule scheme are its *neighbors*. In rules $\vee E$, $\exists E$, the conclusion C forms a *bond* with its premisses, with two bonds for the former rule. (This notion corresponds to Prawitz' "segment.") A single formula is a limiting case of a bond. Two bonds are neighbors if they have formulas that are neighbors. A *hillock* (*Hügel*) is a bond such that its neighboring bonds consist of its subformulas. (These must be proper subformulas or else two bonds become one.) Thus, a hillock is a formula or a sequence of identical formulas of a locally maximal length.

A careful reading of Gentzen's thesis shows that completed derivations in natural deduction do not have open assumptions. Thus, the endformula of a derivation is bound to be the longest formula in a normal derivation. An "inner" hillock is one that is a hillock but not the endformula. Gentzen's proof of normalization considers inner hillocks of maximal formula length, but before the proof can be given, yet another concept has to be introduced. A formula stands *higher* in a derivation than another formula if it stands above the latter. In addition, formulas in the derivations of the minor premisses of rules $\vee E$, $\supset E$, and $\exists E$ stand higher than the main premiss and formulas that stand above the major premiss. A *main hillock* in a derivation is an inner hillock of maximal length such that no other hillock of the same length stands higher than it. Gentzen comes now to the formulation of his "Hügelsatz" or "hillock theorem": A derivation with inner hillocks can be so transformed that one main hillock is removed and the number of main hillocks of maximal length reduced, with no new hillocks of greater length created.

The proof of the hillock theorem has four stages. The first concern is to have distinct eigenvariables, and to ensure that derivations remain correct under substitution of variables. These steps are undertaken with great care. In the second stage, permutative conversions are made to reduce the number of formulas in a bond. The derivations are (*Ms. ULS*, p. 17):

$$\dfrac{\dfrac{\overset{\Gamma}{|C}\quad \Delta}{|C}\,I\ \exists E \qquad \Theta}{\underset{\Lambda}{D}}\,II \qquad\qquad \dfrac{\dfrac{\overset{\Gamma}{|C}\quad \Delta \quad \overset{\Pi}{|C}}{|C}\,I\ \vee E \qquad \Theta}{\underset{\Lambda}{D}}\,II$$

Table 9. Permutation convertibilities in Gentzen's normalization proof.

The order of the premisses differs from the one given earlier in the manuscript. Formulas that form bonds are indicated by $|C$. The upper case Greek letters

denote other parts of the derivations. The transformed derivations are

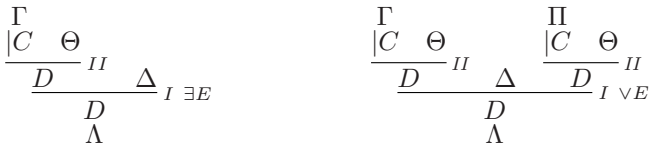


Table 10. The results of permutation conversions.

It is seen that the sequences of bonded formulas are shortened by one in these transformations.

Permutative conversions are repeated until a main hillock consists of just one formula instead of a bonded sequence. Now begins the actual reduction of the derivation. This is divided into two cases according to whether falsity elimination was applied in the main hillock. If not, there are five, by today standard, detour conversion schemes (called reductions by Gentzen). It is remarkable that Gentzen uses the square bracket notation in these, as in Table **8**. In a schematic derivation, $[A]$ indicates any number $n \geqslant 0$ of discharged formulas. In the converted derivations, a similar notation indicates the same number of repeated parts of derivations.

The final part of the proof considers the case that rule $\bot E$ has been applied to conclude a formula C. A local maximum is produced if C is the major premiss of an elimination. Gentzen shows that rule $\bot E$ can be applied to infer the conclusion of the elimination directly. The subformula property of normal derivations is an immediate corollary to the hillock theorem.

Gentzen's preserved manuscripts contain further details about the history of natural deduction that we shall discuss in the last Section.

6 LATER DEVELOPMENTS IN NATURAL DEDUCTION

Gentzen used natural deduction in his early proofs of the consistency of arithmetic: in the original proof of 1935 that remained unpublished, and in a somewhat different form in the published version of 1936. He changed the notation into one that displays the open assumptions $A_1, \ldots, A_n$ of each formula B in a derivation on the same line with that formula, as a *sequent*:

$$A_1, \ldots, A_n \rightarrow B$$

Here the assumptions form a *list* with $n \geqslant 0$. The arrow is not any surrogate implication, but a formal representation of the derivability relation of natural deduction. Now, in contrast to the thesis, assumptions could be left open.

As a particular case of a sequent, we have $A \rightarrow A$ which corresponds to making the assumption A in the standard notation for natural deduction derivations.

Sequents of this form are *initial sequents* and begin derivations. The introduction rules are straightforward, as in the following:

$$\frac{\Gamma \to A \quad \Delta \to B}{\Gamma, \Delta \to A\&B}\,\&I \qquad \frac{\Gamma \to A}{\Gamma \to A \vee B}\,\vee I_1 \qquad \frac{\Gamma \to B}{\Gamma \to A \vee B}\,\vee I_2 \qquad \frac{A, \Gamma \to B}{\Gamma \to A \supset B}\,\supset I$$

$$\frac{\Gamma \to A(y/x)}{\Gamma \to \forall x A}\,\forall I \qquad \frac{\Gamma \to A(t/x)}{\Gamma \to \exists A}\,\exists I$$

Table 11. Introduction rules for natural deduction in sequent calculus style.

Note that the lists of assumptions Γ and Δ are independent in rule $\&I$ and are added up in the conclusion. Rule $\forall I$ has the standard variable restriction. The elimination rules that corresponds to the I-rules are:

$$\frac{\Gamma \to A\&B}{\Gamma \to A}\,\&E_1 \qquad \frac{\Gamma \to A\&B}{\Gamma \to B}\,\&E_2$$

$$\frac{\Gamma \to A \vee B \quad A, \Delta \to C \quad B, \Theta \to C}{\Gamma, \Delta, \Theta \to C}\,\vee E \qquad \frac{\Gamma \to A \supset B \quad \Delta \to A}{\Gamma, \Delta \to B}\,\supset E$$

$$\frac{\Gamma \to \forall x A}{\Gamma \to A(t/x)}\,\forall E \qquad \frac{\Gamma \to \exists x A \quad A(y/x), \Delta \to C}{\Gamma, \Delta \to C}\,\exists E$$

Table 12. Elimination rules for natural deduction in sequent calculus style.

In rule $\exists E$, y is an eigenvariable.

For a good formalization, rules for handling the lists of assumptions have to be made explicit. In Gentzen (1936, 5.2.4), the following "structural modifications" are permitted: to exchange the order of two consecutive assumptions, to add an extra assumption, to contract two identical assumptions into one, and to change a bound variable.

Some confusion has been created by the notation for natural deduction in sequent calculus style. For example, Bernays (1970) calls it a sequent calculus. A look at the rules of the calculus shows that I-rules introduce a connective or quantifier at the right of the sequent arrow, in the *succedent* part. E-rules, in turn, make a connective or quantifier disappear from that same part. As we shall see, the latter makes for a fundamental difference with sequent calculus proper.

At the time when Gentzen worked out his system of natural deduction, Stanislaw Jaskowski was also developing a logical system for reasoning with assumptions. Formulas in derivations are arranged in a linear succession, but Jaskowski's paper (1934) remained fragmentary and without substantial results such as a subformula property. The linear variant of natural deduction is followed in many pedagogical expositions of elementary logic. Gentzen found Jaskowski's work by June 1936, when both were in Münster, and considered its linear arrangement of formulas an improvement, a "liberation from the straitjacket of the tree form," into one that

reflects "the linearity of thinking" (the former from unpublished notes, the latter from Gentzen (1934–35, II.2.2.2).

Gentzen's teacher Paul Bernays visited the Institute for Advanced Study in Princeton in 1935–36. He gave a series of lectures that have been preserved in a mimeographed form, with the title *Logical Calculus*. The larger part of the lectures is devoted to an exposition of Gentzen's system of natural deduction, with the rule of double negation included. The presentation, however, is a great step back from what Bernays could have learned from Gentzen's dissertation and its early manuscript version had he studied them well. No mention is made of intuitionistic logic, and the method of truth tables, in the form of valuations, is offered as a means for deciding deducibility, with the effect that "the performing of deductions becomes quite superfluous" (p. 34). The derivations in natural deduction remain on the informal level of Gentzen's first example, with no clear definition of how derivation trees are to be constructed. Thus, formulas appear under the inference line even if they are assumptions, and there is no way of keeping record of which assumptions are open and which discharged at different points of a derivation. The only use of the calculus is in a proof of equivalence with an axiomatic calculus that is inspired by the rules of natural deduction, say, in the style of the axioms $A\&B \supset A$ and $A\&B \supset B$ in place of the two rules $\&E$. Normalization and the subformula property as well as their applications are not mentioned even if these were the very motives for studying natural deduction in the first place.

On a completely different level is the paper by Ingebrigt Johansson, submitted in 1935. It appears that he had worked extensively on his system of *minimal* logic, a weakening of intuitionistic logic in which $\bot \supset C$ should not be a provable formula. The basis was an axiomatic formulation, but treatments of minimal logic both as a system of natural deduction and as a sequent calculus are also given, in what is the second published work on these calculi altogether. Johansson uses sequent calculus to conclude that the inference to B from $(A\&\neg A) \vee B$ is an *admissible* rule in minimal logic, even if the formula $(A\&\neg A) \vee B \supset B$ remains underivable. The notion of admissibility, even if no German equivalent to such a word is used, is quite clear: A step of inference is admissible in a system if its conclusion is derivable in the system whenever the premisses are.

The system of natural deduction lay mostly dormant for some thirty years, until the thesis of Dag Prawitz of 1965, *Natural Deduction: A Proof-Theoretical Study*. It gives the details of normalization of derivations in natural deduction, also given in a somewhat different and less detailed form in Raggio (1965).

The order in which Prawitz (1965) presented the normalization theorem was different from Gentzen's in *Ms. ULS*, naturally unknown to the former. Prawitz gives first a normalization theorem and subformula property for a system of natural deduction for classical logic. This system contains no disjunction or existence. Therefore, leaving for a while aside the rule that makes the system classical, detour convertibilities could be made directly, without the difficulty provoked by the elimination rules for disjunction and existence. Such direct convertibility was preserved under the addition of a rule of *indirect proof* for atomic formulas P, a

rule that made the system classical:

$$\dfrac{\begin{array}{c}[\neg P]\\ \vdots\\ \bot\end{array}}{P}\,Raa$$

Table 13. Prawitz' rule of indirect proof.

Prawitz showed that in the $\vee, \exists$-free fragment of predicate logic, this rule is admissible for arbitrary formulas A, meaning here that whenever there is a derivation of $\bot$ from the assumption $\neg A$, a derivation of the conclusion A can be constructed. In Gentzen's thesis, various classical rules were suggested, such as the rule of double negation that concludes A from $\neg\neg A$, with A an arbitrary formula and no subformula property. The system was designated *NK*, but Gentzen failed to obtain a normalization theorem for even a fragment of classical logic.

In a second stage, Prawitz considered intuitionistic natural deduction for the full language of predicate logic and reduced its normalization to the deletion of detour convertibilities, through permutation convertibilities that reduced the lengths of the gentzenian "bonds," or "segments" in Prawitz' terminology, until there was just one formula in a segment. The overall structure of the two results, when intuitionistic logic is considered, is the same as that in Gentzen's proof. A direct proof of normalization that combines Prawitz' two results, without the rule of indirect proof of course, is found in the textbook *Basic Proof Theory* by Troelstra and Schwichtenberg. It is, in all essential respects, the same proof as Gentzen's original one.

In the late 1960s, *strong* normalization became an issue: Prawitz (1971), using previous work of William Tait (1967) and Jean-Yves Girard (1971), proved that non-normalities in a derivation can be converted in any order, with a terminating normalization process and a unique normal derivation as a result. Gentzen seems not to have noticed the latter. He states in fact the contrary, no doubt motivated by the failure of this property for the elimination of cuts in a sequent calculus derivation.

At about the same time as strong normalization was studied, the *Curry-Howard* correspondence emerged. Curry had observed in the book on combinatory logic, Curry and Feys (1958), the analogy between implication elimination in natural deduction and functional application. The idea was as old as intuitionistic logic: By the "BHK-explanation" (for Brouwer-Heyting-Kolmogorov), the forms of propositions in intuitionistic logic express prescriptions on how to prove those propositions: A conjunction $A\&B$ is proved by proving A and B separately, a disjunction $A \vee B$ by proving one of A and B, and an implication $A \supset B$ by showing how to convert any proof of A into some proof of B, and so on. These explanations come very close to the introduction rules of natural deduction, but it remains unknown what their effect on Gentzen's thought was. An early, even if rather vague, presentation of intuitionistic logic in these terms is Heyting (1931). In Kolmogorov's

paper (1932), propositions are interpreted as problems, and the BHK-explanations show how solutions to composite problems reduce to their parts.

The Curry-Howard correspondence, from the paper by Howard of 1969, but published only in 1980, is based on the "formulas-as-types" principle, or in another jargon, on the "propositions-as-sets" principle. A proposition is thought of as its set of proofs. Truth of a proposition corresponds to the non-emptiness of the set. An implication $A \supset B$ is identified with the set of functions from A to B. Thus, if $f : A \supset B$ and $a : A$, then functional application gives $f(a) : B$. The reverse, what corresponds to the introduction of an implication, is captured by the principle of functional abstraction of Alonzo Church's λ-calculus (1932, 1941).

The intuitionistic explanation of implication was criticized as being circular in the 1930s, the most prominent critic being Gödel, as in his Zilsel lecture (1938), and in the Yale lecture (1941): The proof of an implication $A \supset B$ is reduced back to a method that converts an arbitrary proof of A into some proof of B. This explanation assumes the notion of an arbitrary proof and cannot therefore function as an explanation of the proof of an implication. Gödel's own response was what came to be known as the Dialectica-interpretation, seen in emergence in the Zilsel lecture and developed fully in the Yale lecture. The interpretation became generally known only with its publication in the journal *Dialectica* in 1958. Today it is seen that the circularity is avoided through a normal form theorem: The explanation refers in the first place to direct proofs of a proposition that have a specific form, and the arbitrary proofs in Heyting's explanation are reduced back to the direct ones that appear through a normalization process. This conceptual justification of intuitionistic logic was worked out by Michael Dummett (1975) and Per Martin-Löf (1975).

The connections between natural deduction, typed λ-calculus, and constructive type theory have led to a proof-theoretical notion of computation. Functional application corresponds to implication elimination: It gives an expression in a non-evaluated form, and the computation of the value of the function is effected by what corresponds exactly to a normalization.

7 SEQUENT CALCULUS

Gentzen's system *NJ* gives intuitionistic logic. As mentioned, Gentzen obtained a system of classical logic *NK* by adding a rule of double negation. Another possibility that he mentions was to let instances of the law of excluded middle appear among the topformulas of derivations. Now not all formulas in a (normal) derivation need be subformulas of the open assumptions or the conclusion. To obtain a proof system for classical logic with the subformula property, Gentzen devised a calculus in which the single conclusion of each rule of natural deduction is replaced by a finite number of *possible cases*. Such cases, say Θ, under the open assumptions, say Γ, are expressed formally by a sequent written as $\Gamma \longrightarrow \Theta$. In Gentzen, Γ and Θ are lists of formulas, so that we can render sequent calculus as "a calculus with lists of assumptions and cases." The German word *Sequenz* means

sequence. The use of the English "sequent" as a noun stems from Kleene's book *Introduction to Metamathematics* (p. 441), the motivation being that "sequence" had another formal use in that book. This neologism has been taken over to several other languages, with such contrived nouns as the Italian "sequente."

We shall first present sequent calculus as it appears from Gentzen's published papers, and then go on in the next Section to discuss its birth as seen from the manuscript version of *Ms. ULS*.

Gentzen's original explanation (1934–35, I.2.4) of the meaning of a sequent $A_1, \ldots, A_m \rightarrow B_1, \ldots, B_n$ was that its content is the same as that of the formula $A_1 \& \ldots \& A_m \supset B_1 \vee \ldots \vee B_n$. A review of that work by Gentzen's contemporary Arnold Schmidt in Göttingen instead contains the following: A sequent $A_1, \ldots, A_m \rightarrow B_1, \ldots, B_n$ means that "B_1 *or* $\ldots B_n$ depends on the assumptions A_1 *and* $\ldots A_m$" (Schmidt 1935, p. 145). It is clear by circumstantial evidence that this suggestion comes from Gentzen.

In reference to his first paper (1936) on the consistency of arithmetic, Gentzen writes in his second paper on the same topic (1938, 1.2):

> In the previous work I had introduced the concept of a sequent, with just one succedent formula, in its immediate connection to the natural representation of mathematical proofs (5). It is possible to arrive at the new, symmetric concept of a sequent also from that same point of view, namely, by striving at a particularly natural representation of the *division into cases* (see 4 of the previous work, and in particular 5.26). Namely, a $\vee$-elimination can now be represented simply as: From $\rightarrow A \vee B$ one concludes $\rightarrow A, B$, to be read as: 'Both possibilities A and B obtain'.

A formal calculus of sequents requires, besides logical rules, also rules for the handling of lists of formulas in the antecedent (left) and succedent (right) parts of sequents. Gentzen gives these first, named the *structural rules* of sequent calculus:

$$\frac{\Gamma \rightarrow \Theta}{D, \Gamma \rightarrow \Theta}\, LW \qquad \frac{\Gamma \rightarrow \Theta}{\Gamma \rightarrow \Theta, D}\, RW$$

$$\frac{D, D, \Gamma \rightarrow \Theta}{D, \Gamma \rightarrow \Theta}\, LC \qquad \frac{\Gamma \rightarrow \Theta, D, D}{\Gamma \rightarrow \Theta, D}\, RC$$

$$\frac{\Delta, D, E, \Gamma \rightarrow \Theta}{\Delta, E, D, \Gamma \rightarrow \Theta}\, LE \qquad \frac{\Gamma \rightarrow \Theta, E, D, \Lambda}{\Gamma \rightarrow \Theta, D, E, \Lambda}\, RE$$

$$\frac{\Gamma \rightarrow \Theta, D \quad D, \Delta \rightarrow \Lambda}{\Gamma, \Delta \rightarrow \Theta, \Lambda}\, Cut$$

Table 14. The structural rules of sequent calculus LK.

Here L and R stand for left and right (A and S in Gentzen), W for weakening (*Verdünnung*, abbr. *Vd*), C for contraction (*Zusammenziehung, Zz*), and E for exchange (*Vertauschung, Vt*). Note the mirror-image arrangement of the rules,

confirmed by the deviation from alphabetical order in the premiss of rule *RE*. Rule *Cut* (*Schnitt*) is self-symmetric. It permits the *composition* of two derivations, one with the case D, the other with the assumption D.

In the following logical rules of sequent calculus, the formulas displayed in the premisses are the *active* formulas of a rule (*Nebenformeln*), and the one with the connective or quantifier the *principal* formula (*Hauptformel*). The upper case Greek letters indicate what are today called the *contexts* of the rules; these can be empty. In general, the antecedent or the succedent of a sequent can be empty. If both are empty, we have the *empty sequent* $\rightarrow$. The assumption of a formula A in natural deduction corresponds to the initial sequent $A \rightarrow A$ by which derivations start.

The logical rules of Gentzen's classical sequent calculus are:

$$\frac{\Gamma \rightarrow \Theta, A \quad \Gamma \rightarrow \Theta, B}{\Gamma \rightarrow \Theta, A\&B}\, R\& \qquad \frac{A, \Gamma \rightarrow \Theta \quad B, \Gamma \rightarrow \Theta}{A \vee B, \Gamma \rightarrow \Theta}\, L\vee$$

$$\frac{A, \Gamma \rightarrow \Theta}{A\&B, \Gamma \rightarrow \Theta}\, L\&_1 \qquad \frac{B, \Gamma \rightarrow \Theta}{A\&B, \Gamma \rightarrow \Theta}\, L\&_2 \qquad \frac{\Gamma \rightarrow \Theta, A}{\Gamma \rightarrow \Theta, A \vee B}\, R\vee_1 \qquad \frac{\Gamma \rightarrow \Theta, B}{\Gamma \rightarrow \Theta, A \vee B}\, R\vee_2$$

$$\frac{\Gamma \rightarrow \Theta, A(y/x)}{\Gamma \rightarrow \Theta, \forall x A}\, R\forall \qquad \frac{A(y/x), \Gamma \rightarrow \Theta}{\exists x A, \Gamma \rightarrow \Theta}\, L\exists$$

$$\frac{A(t/x), \Gamma \rightarrow \Theta}{\forall x A, \Gamma \rightarrow \Theta}\, L\forall \qquad \frac{\Gamma \rightarrow \Theta, A(t/x)}{\Gamma \rightarrow \Theta, \exists x A}\, R\exists$$

$$\frac{A, \Gamma \rightarrow \Theta}{\Gamma \rightarrow \Theta, \neg A}\, R\neg \qquad \frac{\Gamma \rightarrow \Theta, A}{\neg A, \Gamma \rightarrow \Theta}\, L\neg$$

$$\frac{A, \Gamma \rightarrow \Theta, B}{\Gamma \rightarrow \Theta, A \supset B}\, R\supset \qquad \frac{\Gamma \rightarrow \Theta, A \quad B, \Delta \rightarrow \Lambda}{A \supset B, \Gamma, \Delta \rightarrow \Theta, \Lambda}\, L\supset$$

Table 15. The logical rules of sequent calculus LK.

Rules $R\forall$ and $L\exists$ have the variable restriction that the eigenvariable y must not occur free in the conclusion. The rules are displayed in a mirror-image symmetry down to the last two that have no symmetry. These are separated in the German original by a horizontal line, the ones above by a vertical line. The point is to display the duality of & and $\vee$ resp. $\forall$ and $\exists$, and the self-duality of $\neg$ (1934–35, III.2.2.4). Note that the symmetric arrangement is not reproduced in the layout of the 1969 English translation of Gentzen's papers.

Gentzen's system of intuitionistic sequent calculus *LJ* is obtained from the classical calculus *LK* as a special case, by the requirement that the succedent consist of at most one formula. This feature is one of Gentzen's essential design principles in his putting up the calculus *LK*, and it explains some of the main aspects of the logical rules. First of all, rule $L\supset$ has single-succedent instances only if the contexts in the two premisses are independent of each other. Note that the other two-premiss rules have instead "shared," identical contexts. The explanation of

this feature is that the proof of the main result for sequent calculus, Gentzen's famous *Hauptsatz*, is simplified a bit.

To display a further special feature of the calculus LK, consider the steps of derivation

$$\dfrac{\dfrac{\dfrac{\dfrac{A, B, \Gamma \to \Theta}{A\&B, B, \Gamma \to \Theta}\, L\&_1}{B, A\&B, \Gamma \to \Theta}\, LE}{A\&B, A\&B, \Gamma \to \Theta}\, L\&_2}{A\&B, \Gamma \to \Theta}\, LC$$

The result would be obtained in one step from the topsequent if instead of Gentzen's two rules one used a single left conjunction rule that concludes $A\&B, \Gamma \to \Theta$ from $A, B, \Gamma \to \Theta$. The above steps of derivation show that this rule is a *derivable* rule in LK. In the other direction, if say $A, \Gamma \to \Theta$ is given, a weakening with B followed by an exchange and the modified rule of left conjunction gives $A\&B, \Gamma \to \Theta$, and similarly if $B, \Gamma \to \Theta$ is given. Thus, Gentzen's rules are derivable by the modified rule.

The presence of two left conjunction rules is explained by Gentzen's wish to display the duality of & and $\vee$: A dual single right disjunction rule would conclude $\Gamma \to \Theta, A \vee B$ from $\Gamma \to \Theta, A, B$, but this rule has no single-succedent instances. The intuitionistic calculus would not come out as a special case, so Gentzen's solution was to have two left conjunction rules.

Gentzen's remark in the end of the passage from 1938 quoted above amounts to the *invertibility* of a single right disjunction rule, one that concludes $\Gamma \to \Theta, A \vee B$ from $\Gamma \to \Theta, A, B$. Invertibility of logical rules as a general phenomenon was discovered by Ketonen (1944) and we shall discuss it shortly. An anticipation in *Ms. ULS* is discussed in Section 8.

It should be clear that one can arrive at such choices as Gentzen's regarding the details of a logical calculus only by trying out various formulations. What they were and how Gentzen reasoned about them is difficult to say, for whatever papers and notes Gentzen had left behind in Göttingen, they were destroyed in 1964 after the death of Arnold Schmidt together with Schmidt's papers. Essential new light on the discovery of sequent calculus is now shed by the manuscript version of Gentzen's thesis.

Some variants of sequent calculi can be found in Gentzen's papers. He mentions in (1934–35, III.2.2) the possibility to let, among others, sequents of the form $A, A \supset B \to B$ be among the topsequents of derivations, and uses this in (1936) and (1943) (the former natural deduction in sequent calculus style). Given the premisses of $L\supset$, so $\Gamma \to \Theta, A$ and $B, \Delta \to \Lambda$, the conclusion is obtained as follows:

$$\dfrac{\dfrac{\dfrac{\Gamma \to \Theta, A \quad A, A \supset B \to B}{\Gamma, A \supset B \to \Theta, B}\, Cut \qquad B, \Delta \to \Lambda}{\Gamma, A \supset B, \Delta \to \Theta, \Lambda}\, Cut}{A \supset B, \Gamma, \Delta \to \Theta, \Lambda}\, LE^*$$

The asterisk indicates as many left exchanges as are needed to bring $A \supset B$ in the head of the antecedent. The two cuts are innocuous, because both cut formulas are subformulas of $A \supset B$ that appears in the conclusion.

In Gentzen (1938), implication is considered a defined notion. This possibility is mentioned also in (1934–35, III.2.4.1), with the remark that the definition can be given only in the classical calculus *NK*. In the sequent calculi of (1934–35), all the standard connectives and quantifiers were present. A sufficient reason for this is the requirement that the intuitionistic calculus be a special case of the classical one. Possibly Gentzen had also some other ideas in the matter, because he states that "not all of the rules are mutually independent, i.e., one could substitute some by others. However, if we left some out, the 'Hauptsatz' would not hold anymore." (1934–35, III.2.1). If we take just some fragment of the language of predicate logic, the rules for the connectives and quantifiers are complete for that fragment. It seems that Gentzen's thought was the following: If one defines logical operations by others through equivalences, as in $A \vee B \supset\subset \neg(\neg A \& \neg B)$, these equivalences can be put into use only by using them as cut formulas.

We now come to Gentzen's *Hauptsatz*: The rule of cut can be eliminated from derivations by permuting it up until it reaches initial sequents and gets removed as superfluous. A mere look at the rest of the rules shows that derivations free of the cut rule have the subformula property.

The proof of cut elimination proceeds by considering an uppermost cut, as in a typical case of

$$\dfrac{\dfrac{\Gamma \to \Theta, A \quad \Gamma \to \Theta, B}{\Gamma \to \Theta, A\&B}\, R\& \qquad \dfrac{A, \Delta \to \Lambda}{A\&B, \Delta \to \Lambda}\, L\&_1}{\Gamma, \Delta \to \Theta, \Lambda}\, Cut$$

Cut is permuted up as follows:

$$\dfrac{\Gamma \to \Theta, A \quad A, \Delta \to \Lambda}{\Gamma, \Delta \to \Theta, \Lambda}\, Cut$$

Cut on $A\&B$ is replaced by a cut, or two cuts if $R\supset$ and $L\supset$ were applied, on *strictly shorter* formulas. We note in passing that the "shared" contexts of rule $R\&$ do a work here, as mentioned by Gentzen: Had they been independent, there would have to be steps of weakening that restore the original contexts of the conclusion of cut.

Another type of cut is met when the cut formula is not principal in the left premiss, as in

$$\dfrac{\dfrac{C, \Gamma \to \Theta, A\&B}{C\&D, \Gamma \to \Theta, A\&B}\, L\&_1 \qquad \dfrac{A, \Delta \to \Lambda}{A\&B, \Delta \to \Lambda}\, L\&_1}{C\&D, \Gamma, \Delta \to \Theta, \Lambda}\, Cut$$

Cut is permuted up as follows:

$$\dfrac{\dfrac{C, \Gamma \rightarrow \Theta, A \& B \quad \dfrac{A, \Delta \rightarrow \Lambda}{A \& B, \Delta \rightarrow \Lambda}\, L\&_1}{C, \Gamma, \Delta \rightarrow \Theta, \Lambda}\, Cut}{C \& D, \Gamma, \Delta \rightarrow \Theta, \Lambda}\, L\&_1$$

Here the *rank* of the cut is diminished by one: Trace up the cut formula from both premisses of cut until its uppermost appearance. Rank is defined as the sum of the number of steps of derivation from these first appearances to the cut. In Gentzen's proof of cut elimination, an uppermost cut in a derivation tree is eliminated, so the proof is an induction on the number of cuts along branches of a derivation. Secondly one considers the length of the cut formula, and thirdly the rank of the cut.

When a cut is permuted up, it eventually reaches an initial sequent. When one premiss of a cut is an initial sequent, the conclusion is identical to the other premiss and the cut is deleted.

Gentzen met with one problematic case of cut elimination, one in which the cut formula in the right premiss is principal in a contraction:

$$\dfrac{\Gamma \rightarrow \Theta, D \quad \dfrac{D, D, \Delta \rightarrow \Lambda}{D, \Delta \rightarrow \Lambda}\, LC}{\Gamma, \Delta \rightarrow \Theta, \Lambda}\, Cut$$

Two cuts on D are needed, with the result

$$\dfrac{\dfrac{\Gamma \rightarrow \Theta, D \quad \dfrac{D, D, \Delta \rightarrow \Lambda}{\Gamma, D, \Delta \rightarrow \Theta, \Lambda}\, Cut}{D, \Gamma, \Delta \rightarrow \Theta, \Lambda}\, LE^*}{\Gamma, \Gamma, \Delta \rightarrow \Theta, \Theta, \Lambda}\, Cut}{\Gamma, \Delta \rightarrow \Theta, \Lambda}\, LC^*, RC^*$$

Here LE^* indicates as many left exchanges as are needed to bring D in front of Γ and LC^*, RC^* left and right contractions that eliminate the duplications of Γ and Θ. The cut formula remains the same, but the rank of the lower cut is not reduced, even if the left exchanges LE^* were absent. Gentzen's solution was to introduce a stronger form of cut, in which any number of cut formulas can be eliminated from the premisses of cut. He called it *Mischung*, rendered as *mix* in some older English literature, and today often called *multicut*:

$$\dfrac{\Gamma \rightarrow \Theta \quad \Delta \rightarrow \Lambda}{\Gamma, \Delta^* \rightarrow \Theta^*, \Lambda}\, Cut^*$$

Table 16. Gentzen's mix rule.

In the rule, Θ and Δ are lists of formulas in which a "mix formula" D appears at least once in both, and Θ^* and Δ^* are lists obtained from Θ and Δ through the

deletion of all occurrences of D. Gentzen writes that the new rule is introduced "to make the proof easier" (1934–35, III.3.3.1). Intrigued by this remark, the present author produced a proof of Gentzen's *Hauptsatz* without multicut, by an analysis of how the contraction formula first appeared in the derivation of the right premiss of cut (von Plato 2001).

Cut is a special case of multicut, so that the elimination theorem for the latter gives also an elimination theorem for the former.

The mix rule treats the antecedents and succedents of sequents without regard to the order of formulas. This feature can be carried through systematically for all the rules of sequent calculus by using multisets of formulas instead of lists. Then the exchange rules can be dropped out. If one treated the standard rule of cut in this multiset way, the left exchanges LE^* in the problematic case of permutation of cut given above would disappear, but still the lower cut would not have a reduced rank.

The idea has been often tried of treating the antecedents and succedents as sets instead of lists or multisets, which would lead to the exchange and contraction rules being dropped out. Gentzen does not mention this possibility, and the conclusion cannot be other than that he thought it not worth mentioning. The problem is the purely extensional notion of a set: Two sets are equal if they have the same members, and equals should be substitutable by equals. Logical rules, instead, are intensional, they work on the forms of expressions. As an example, consider rule $R\&$, and let the premisses be $A, B \to C, D$ and $A, B \to C, E$. (It is understood that $A, B \to C, D$ stands for $\{A, B\} \to \{C, D\}$, and so on.) Now $R\&$ gives $A, B \to C, D\&E$ as a conclusion. If, however, C and D happen to be identical, which is not in any way excluded by the schematic rule, then the premisses, with sets rather than lists or multisets, are $A, B \to C$ (by the identity $\{C, D\} = \{C\}$) and $A, B \to C, E$, and rule $R\&$ does not match the form of these two sequents. Attempts at circumventing this kind of a problem lead invariably to rules that do not operate on sets, but on expressions for sets, with some rule of substitution that tells when one expression for a set can be replaced by another expression for the same set. In the end, nothing at all is gained by the exercise: A rule that permits to substitute an expression $\{C, C, D, \ldots\}$ for a set by the expression $\{C, D, \ldots\}$ for an equal set is a rule of contraction.

The first applications of the cut elimination theorem, to intuitionistic propositional logic, were listed above. Another application in Gentzen (1934–35) was to the consistency problem of arithmetic. If the induction scheme is left out, cut elimination goes through and the underivability of the empty sequent $\to$, expressing inconsistency, follows. This consistency result was known from the work of Johann von Neumann and Jacques Herbrand, the names that Gentzen mentions. For him, this application showed the way to a future consistency proof of the whole of Peano arithmetic within the new proof-theoretical system of sequent calculus.

8 THE ORIGINS OF SEQUENT CALCULUS

The manuscript version of Gentzen's thesis contains the first steps in sequent calculus. The pertinent pages of *Ms. ULS* are numbered 29–32, but these numbers are later additions to a text written in the "deutsche Schrift" of section I of *Ms. ULS*, discussed in Section 5. A calculus *NL3* is introduced, described as a "natural-logistic parallel calculus." The basic expressions of the calculus are taken over from Hertz and are called "sentences" (*Satz*). They are written

$$A_1, \ldots, A_\nu \supset B$$

and the reading is: "Proof of a proposition B from the assumptions $A_1, \ldots, A_\nu$." The notion of a proof is defined inductively, with the "natural" calculus at left, the "parallel" calculus at right:

I. A is a proof of A from A.

$$A \qquad\qquad A \supset A$$

Next a first group of single proof steps is translated into sentences:

1.) The proof of Ph from $(x)Px$ becomes

$$\frac{(x)Px}{Ph} \qquad\qquad (x)Px \supset Ph$$

Similarly, the rest of the rules become:

$$\frac{Ph}{(Ex)Px} \qquad Ph \supset (Ex)Px$$

$$\frac{A\&B}{A} \quad \frac{A\&B}{B} \qquad A\&B \supset A \qquad A\&B \supset B$$

$$\frac{A}{A \vee B} \quad \frac{B}{A \vee B} \qquad A \supset A \vee B \qquad B \supset A \vee B$$

$$\frac{A \quad A \to B}{B} \qquad A, A \to B \supset B$$

2.) A second group of proof steps come from the schematic rules of natural deduction that are translated into proper rules of *NL3*. The first rule is $\forall I$. Gentzen writes that "from a proof of P from $A_1 \ldots A_\nu$ emerges (through the addition of $(x)Px$, below) a proof for $(x)Px$ from $A_1 \ldots A_\nu$." The rules, in "parallel" writing are, with Γ abbreviating $A_1 \ldots A_\nu$:

$$\frac{\dfrac{A_1 \ldots A_\nu}{A_1 \ldots A_\nu}}{\begin{array}{c}\vdots\\ Pa\end{array}}{(x)Px} \qquad\qquad \frac{\Gamma \supset Pa}{\Gamma \supset (x)Px}$$

It is not explained why the A_i are repeated under the inference line. The part from $A_1 \ldots A_\nu$ to Pa at the left side, with a the eigenvariable, is circled around and becomes the premiss of the rule at right.

The first left rule comes from $\exists E$, and Gentzen writes that "from a proof of B from $\Gamma, [Pa]$ emerges a proof of B from $\Gamma, (Ex)Px$." Brackets denote any number $\geqslant 0$ of occurrences of Pa, with a the eigenvariable. The parallel writing continues:

$$\dfrac{\dfrac{\Gamma\ (Ex)Px}{\Gamma\ [Pa]}\ \vdots\ B}{B} \qquad\qquad \dfrac{\Gamma\ [Pa] \supset B}{\Gamma\ (Ex)Px \supset B}$$

The part from $\Gamma\ [Pa]$ to B at left is again circled.

Rule $\&I$ is the first one with two premisses:

$$\dfrac{\dfrac{\Gamma}{\Gamma}\ \vdots\ A \quad \dfrac{\Delta}{\Delta}\ \vdots\ B}{A\&B} \qquad\qquad \dfrac{\Gamma \supset A \quad \Delta \supset B}{\Gamma\ \Delta \supset A\&B}$$

Note that the contexts are independent, as in natural deduction. Rules $\vee E$ and $\supset I$ are:

$$\dfrac{\Gamma \quad A \vee B \quad \Delta}{[A]\Gamma \quad [B]\Delta}\ \vdots\ \dfrac{C \quad C}{C} \qquad\qquad \dfrac{[A]\Gamma \supset C \quad [B]\Delta \supset C}{(A \vee B)\Gamma\ \Delta \supset C}$$

$$\dfrac{\dfrac{\Gamma}{[A]\Gamma}\ \vdots\ B}{A \to B} \qquad\qquad \dfrac{[A]\Gamma \supset C}{\Gamma \supset A \to B}$$

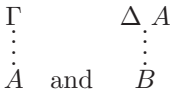
Finally, the rule of induction is translated into one concluding $\Gamma\ P1 \supset Ph$ from $\Gamma\ [Pa] \supset Pa'$.

Next comes stage II of the inductive definition of derivations:

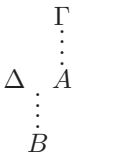
II. Addition of a condition (Thinning)

$$\begin{array}{c}\Gamma\\ \vdots\\ B\end{array} \quad \text{becomes} \quad \begin{array}{c}\Gamma\ A\\ \vdots\\ B\end{array} \qquad\qquad \dfrac{\Gamma \supset B}{\Gamma\ A \supset B}$$

Combination (Cut)



become



$$\frac{\Gamma \supset A \quad \Delta\ A \supset B}{\Gamma\ \Delta \supset B}$$

Gentzen describes his "natural-logistic parallel calculus" *NL3* as follows (*Ms. ULS*, p. 31):

> The *logistic form of proof* arises from the natural one as follows: The parts of proofs that arise in succession (*nacheinander*) in the construction of a natural proof are written symbolically one below the other (*untereinander*).

The structure of proofs is described to have a tree form, with "axioms" as leaves. These are of the forms

$$A \supset A, \quad (x)Px \supset Ph, \quad Ph \supset (Ex)Px, \quad A\&B \supset A, \quad A\&B \supset B,$$

$$A \supset A \vee B, \quad B \supset A \vee B, \quad A, A \rightarrow B \supset B.$$

Table 17. Axioms of Gentzen's "parallel calculus" NL3.

(The last axiom has a comma in the original, whereas Gentzen otherwise just added some extra space.) Axioms are followed by applications of the rules, cut included.

Gentzen notes that each logical operation has an axiom (or a pair of axioms) and a rule, and that rules such as $\&I$ could be replaced by corresponding axioms, in this case the axiom $A, B \supset A\&B$. This would, however, brake down a symmetric arrangement. In the other direction, all axioms for the "logical signs" can be replaced by the following rules (*Ms. ULS*, p. 32):

$$\frac{\Gamma \supset (x)Px}{\Gamma \supset Ph} \qquad \frac{\Gamma\ (Ex)Px \supset C}{\Gamma\ Ph \supset C} \qquad \frac{\Gamma \supset A\&B}{\Gamma \supset A} \qquad \frac{\Gamma \supset A\&B}{\Gamma \supset B}$$

$$\frac{A \vee B\ \Gamma \supset C}{A\ \Gamma \supset C} \qquad \frac{A \vee B\ \Gamma \supset C}{B\ \Gamma \supset C} \qquad \frac{\Gamma \supset A \rightarrow B}{A\ \Gamma \supset B}$$

Table 18. Rules that replace axioms in NL3.

As an example, Gentzen gives the rule for the universal quantifier. The axiom $(x)Px \supset Ph$ gives the corresponding rule by

$$\frac{\Gamma \supset (x)Px \quad (x)Px \supset Ph}{\Gamma \supset Ph}$$

The right premiss is indicated as an "axiom" and the step of inference as "cut." In the other direction, the rule gives

$$\frac{(x)Px \supset (x)Px}{(x)Px \supset Ph}$$

The premiss here is indicated as a "structural axiom." The above rules are, writes Gentzen, "largely analogous (inversions) to those in *2*."

The square brackets in the premisses of the logical rules of Gentzen's first sequent calculus *NL3* indicate any number $\geqslant 0$ of copies of active formulas of sequent calculus rules. Correspondingly, there is no rule of contraction. (By the same, with zero occurences one could do away with weakening, though only for compound formulas.) The possibility of permitting any number of active formulas and no explicit rule of contraction (nor weakening) was (re)discovered by Sara Negri in her studies of sequent calculi for linear logic. It was worked out systematically for intuitionistic and classical sequent calculi in the paper Negri and von Plato (2001a), titled *Sequent calculus in natural deduction style*. Little did we know that we had (re)developed the very first form of sequent calculus ever!

The *Ms. ULS* contains also a brief consideration of a classical "logistic-dualistic" sequent calculus *LDK*. The sentences have the form, now with commas,

$$A_1, \ldots, A_\mu \supset B_1, \ldots, B_\nu$$

The meaning of these sentences is explained as in the thesis: If each of the A holds, then at least one of the B holds. It is further stated that "the order and multiplicity of the A, resp. B, makes no difference."

System *LDK* consists of the "structural axiom" $A \supset A$ and the related structural rules of weakening and cut, written as:

$$\frac{\Gamma \supset \Delta}{\Theta\Gamma \supset \Delta\Lambda} \qquad\qquad \frac{\Gamma \supset \Delta A \quad A\Theta \supset \Lambda}{\Gamma\Theta \supset \Delta\Lambda}$$

Table 19. Weakening and cut in the "dualistic" calculus LDK.

The axioms for the logical signs are given next, with implication a defined notion. The arrangement displays the dualities, just as in the structural and logical rules of the published thesis.

Axioms

&	$\vee$
$A\&B \supset A$	$A \supset A \vee B$

$A\&B \supset B$	$B \supset A \vee B$
$A, B \supset A\&B$	$A \vee B \supset A, B$
	$\neg$
$A, \neg A \supset$	$\supset A, \neg A$
$(\ )$	(E)
$(x)Px \supset Ph$	$Ph \supset (Ex)Px$
	Rules
$\dfrac{\Gamma \supset Pa}{\Gamma \supset (x)Px}$	$\dfrac{Pa \supset \Gamma}{(Ex)Px \supset \Gamma}$

Table 20. Logical axioms and rules of the calculus LDK.

Here the text abruptly ends.

9 RELATIONS BETWEEN NATURAL DEDUCTION AND SEQUENT CALCULUS

The last chapter of Gentzen's thesis contains proofs of the equivalence of axiomatic, natural deduction, and sequent calculus formulations of logic through a cycle of translations of derivations. The equivalence of natural deduction and axiomatic presentations of logic was treated already in the *Ms. ULS*, section II. The correspondence between natural deduction and sequent calculus is also present in the manuscript, as we saw in the previous Section.

We shall consider here Gentzen's translation from natural deduction to the intuitionistic sequent calculus *LJ*. It consists of two phases, in the first of which each formula C in a derivation is replaced by a sequent $\Gamma \longrightarrow C$ in which Γ lists the open assumptions C depends on. If there was an *I*-rule in the natural deduction derivation, we have, say,

$$\dfrac{\begin{array}{cc}\Gamma & \Delta \\ \vdots & \vdots \\ A & B\end{array}}{A\&B}\,{\scriptstyle \&I} \quad \rightsquigarrow \quad \dfrac{\Gamma \longrightarrow A \quad \Delta \longrightarrow B}{\Gamma, \Delta \longrightarrow A\&B}$$

The translation gives the step $R\&$, modulo a second phase of weakenings in the premisses to get a rule with the shared context Γ, Δ in the premisses. The other *I*-rules translate similarly, and there is a direct correspondence between introductions in natural deduction and right rules in sequent calculus.

With elimination rules, the translation is different. For example, $\supset E$ is translated in the first phase as

$$\dfrac{\begin{array}{cc}\Gamma & \Delta \\ \vdots & \vdots \\ A \supset B & A\end{array}}{B}\,{\scriptstyle \supset E} \quad \rightsquigarrow \quad \dfrac{\Gamma \longrightarrow A \supset B \quad \Delta \longrightarrow A}{\Gamma, \Delta \longrightarrow B}$$

In the next phase a cut is inserted as follows:

$$\dfrac{\Gamma \to A \supset B \qquad \dfrac{\Delta \to A \quad B \to B}{A \supset B, \Delta \to B}\, L\supset}{\Gamma, \Delta \to B}\, Cut$$

The derivation in natural deduction could very well have been normal in Gentzen's sense, i.e., without consecutive *I*-*E*-pairs of rules, yet the result of translation is a sequent calculus derivation with a cut. Gentzen does not comment on this issue. However, another translation relates to the role of the structural rules of sequent calculus: If in rule $\supset I$ several, say two, or no copies of formula A were discharged, the translation gives, respectively,

$$\dfrac{\dfrac{A, A, \Gamma \to B}{A, \Gamma \to B}\, LC}{\Gamma \to A \supset B}\, R\supset \qquad \dfrac{\dfrac{\Gamma \to B}{A, \Gamma \to B}\, LW}{\Gamma \to A \supset B}\, R\supset$$

Thus, the contraction stems from a *multiple* discharge of an assumption in natural deduction, and the weakening from a *vacuous* discharge. The same holds for the translation of rules $\vee E$ and $\exists E$.

The correspondence that Gentzen's work gives between weakening and vacuous discharge and contraction and multiple discharge goes only in one direction, from natural deduction to sequent calculus. It is possible to make the correspondence exact in both directions by changing the rest of the elimination rules so that they have an arbitrary consequence like $\vee E$ and $\exists E$ (as is done in von Plato 2001a). By this change, the cuts produced by Gentzen's translation of rules $\& E$, $\supset E$, and $\forall E$ find an explanation (see von Plato 2003). Furthermore, cuts in which the cut formula is principal in a logical rule in the right premiss of cut, so in a left logical rule, can be interpreted as non-normal instances of the corresponding elimination rule of natural deduction.

A whole branch of logical studies has been inspired by the special role of the structural rules of Gentzen's sequent calculi, so-called "substructural logics." Gentzen himself did not think there could be any meaning in the structural rules in isolation of a logical calculus. In his (1936, 5.244), he called these rules "structural modifications" and stated that "all these possibilities of modification are of a purely *formal nature*. It is only because of special features of the formalism that these rules must be expressly given."

In his thesis, Gentzen stated about the elimination rules that "it should be possible to determine the elimination rules as unique functions of the corresponding introduction rules, on the basis of certain conditions" (1934–35, III.5.12), but did not pursue the matter further. A minimum condition is given by what is called the Gentzen-Prawitz *inversion principle*: The elimination rule of a connective or quantifier should bring back that which is included in the sufficient conditions for introducing that connective or quantifier. The detour conversion schemes have been seen as a formal manifestation of this idea: They justify the elimination rules

in terms of the introduction rules, by showing how the grounds for introducing a formula are recovered in the conversions.

The Gentzen-Prawitz inversion principle does not meet Gentzen's requirement of actually determining the elimination rules from the introduction rules, instead of only justifying them. The principle can be generalized, as in Negri and von Plato (2001, p. 6), into one in which one looks at the arbitrary consequences of the sufficient grounds for introducing a formula, instead of those grounds. For conjunction the grounds are A and B separately, and their arbitrary consequences give the following *general* elimination rule, first found by Schroeder-Heister (1984):

$$\frac{A\&B \qquad \begin{array}{c}[A,B]\\ \vdots\\ C\end{array}}{C}\,\&E$$

Table 21. Schroeder-Heister's conjunction elimination rule.

For implication, the sufficient ground for introducing $A \supset B$ is, in Gentzen's words, "the existence of a derivation of B from A" (1934–35, II.5.13). First-order logic is not able to represent formally within its language the existence of a derivation. Therefore Schroeder-Heister (1984) considered a system of higher-order rules. At about the same time, Martin-Löf (1984) formulated a general scheme for elimination rules in his constructive type theory in which the existence of a derivation can be expressed.

What has been said of implication goes also for universal quantification: The sufficient ground for concluding $\forall x A$ is the existence of a derivation of $A(y/x)$ for an *arbitrary* y. Type theory can hypothesize the existence of a function that produces, for any value of y, a proof of $A(y/x)$, and express that a formula C follows from the existence of such a function.

Looking finally at the rules of natural deduction in sequent calculus style, tables **11** and **12** above, we note that a normal instance of $\vee E$ is

$$\frac{A \vee B \rightarrow A \vee B \quad A, \Delta \rightarrow C \quad B, \Theta \rightarrow C}{A \vee B, \Delta, \Theta \rightarrow C}\,\vee E$$

If the major premiss $A \vee B \rightarrow A \vee B$ is left unwritten, a left disjunction rule of sequent calculus, with independent contexts, is obtained. The same observation can be made about all the elimination rules when given in their general form (as in von Plato 2001a). Thus, with this change of E-rules, normal derivations in natural deduction in sequent calculus style and cut-free derivations in sequent calculus itself become notational variants of each other. The real difference between the two types of calculi becomes evident: The class of derivations with the logical rules of natural deduction is closed with respect to composition of derivations, whereas the class of derivations with the logical rules of sequent calculus is not: A non-normal instance of a logical rule in the former is represented as a left rule and a cut in the latter.

10 LATER DEVELOPMENTS AND APPLICATIONS OF SEQUENT CALCULUS

Gentzen's doctoral thesis marked the birth of structural proof theory, as contrasted to the old axiomatic proof theory of Hilbert. As mentioned, the first one to have used Gentzen's methods in print was Ingebrigt Johansson (1936).

In 1938, Katudi Ono published a paper that used sequent calculus for the study of consistency problems of axiomatic systems. Gentzen himself wrote a review of this paper, reprinted in Menzler-Trott (2001, pp. 91–92).

A remarkable step ahead in the development of systems of sequent calculus was taken by Oiva Ketonen in his doctoral thesis of 1944. Ketonen, a student of mathematics and philosophy in Helsinki, went to Göttingen in 1938 to study proof theory with Gentzen, the closest to a student that the latter may have had. The connection seems to have been established by his philosophy professor Eino Kaila who had met Gentzen in 1936 in Münster. Ketonen recollected later to the present author that Gentzen was a person who "did not say much anything." In a letter of December 1938 to his mother he mentions, in a rather frustrated tone, that finally Gentzen has given him a problem to work on, but unfortunately the letter does not tell what problem it was. If we had to guess, it would be the problem of minimizing the number of eigenvariables in sequent calculus derivations, a problem that was solved by Ketonen's improvement on Gentzen's midsequent theorem.

Ketonen's best-known discovery is a sequent calculus for classical propositional logic the logical rules of which are all invertible. The rules that are modified, in comparison to Gentzen's calculus LK, are as follows:

$$\frac{A, B, \Gamma \to \Theta}{A\&B, \Gamma \to \Theta}\, L\& \qquad \frac{\Gamma \to \Theta, A, B}{\Gamma \to \Theta, A \vee B}\, R\vee \qquad \frac{\Gamma \to \Theta, A \quad B, \Gamma \to \Theta}{A \supset B, \Gamma \to \Theta}\, L\supset$$

Table 22. Classical sequent calculus with invertible rules.

Each connective has just one left and right rule. Further, all two-premiss rules have shared contexts, and Ketonen obtains a fully invertible classical propositional sequent calculus. Invertibility is proved after a proof of cut elimination, in the manner of the following example of the invertibility of rule $R\vee$: Assume the sequent $\Gamma \to \Theta, A \vee B$ to be derivable. Then also $\Gamma \to \Theta, A, B$ is derivable:

$$\dfrac{\Gamma \to \Theta, A \vee B \qquad \dfrac{\dfrac{A \to A}{A \to A, B}\, RW \qquad \dfrac{\dfrac{B \to B}{B \to B, A}\, RW}{B \to A, B}\, RE}{A \vee B \to A, B}\, L\vee}{\Gamma \to \Theta, A, B}\, Cut$$

Derivations in the invertible calculus are found by decomposing the endsequent in any order whatsoever: Given a sequent $\Gamma \to \Theta$ to be derived, choose from Γ or Θ any formula with a connective. The corresponding rule determines uniquely the premisses. Repeating this root-first proof search, as it is often called, formulas

are decomposed into parts until there is nothing to decompose. At this stage, it can be determined if the original sequent $\Gamma \rightarrow \Theta$ is derivable, by controlling if all topsequents are initial sequents. The process terminates in the case of classical propositional logic in a bounded number of steps as determined by the number of connectives in the given sequent.

Ketonen's proof of invertibility of the logical rules of his sequent calculus used the structural rule of cut. Later Kurt Schütte (1950) and Haskell Curry (1963) gave direct proofs of invertibility, the latter with the explicit result that the inversions are *height preserving*: If a given sequent is derivable in at most n steps, the premisses in a rule that can conclude that sequent also have a derivation in at most n steps.

The second part of Ketonen's thesis contained an improvement on Gentzen's *midsequent* theorem. This result, from Gentzen (1934–35, IV.2), states that if a sequent $\Gamma \rightarrow \Theta$ is derivable in *LK* and all quantified formulas are in prenex normal form, there is a "midsequent" $\Gamma' \rightarrow \Theta'$ in the derivation such that all logical steps above the midsequent are propositional, and those below it quantificational. Ketonen shows through suitable permutabilities of rules that the number of quantificational inferences with an eigenvariable, i.e., of instances of $R\forall$ and $L\exists$, can be minimized, with the result that there is a *weakest possible* midsequent, one with the property that if any midsequent is derivable, a weakest one is. He was further able to define a recursively enumerable sequence of all possible midsequents $M_1, M_2, \ldots$ for a sequent $\Gamma \rightarrow \Theta$, by which the undecidability of predicate logic could be seen in a proper light: The derivability of any given midsequent M_i of the sequence $M_1, M_2, \ldots$ is decidable, but at no finite stage can it be known if a midsequent $\Gamma' \rightarrow \Theta'$ for the sequent $\Gamma \rightarrow \Theta$ to be derived shows up.

Gentzen (1934–35, IV.2.2) notes that Herbrand's theorem is (for formulas in prenex form) a special case of his midsequent theorem. He used the midsequent theorem for a proof of the consistency of arithmetic without the induction principle (1934–35, IV.3).

Ketonen's main aim in defining his weakest possible midsequents was to use them in proofs of underivability. He writes in the introduction to his final, third chapter that "in proof theory, the most common task – the exclusive one in the applications we discuss – is to show the underivability of a given sequent. Our method is especially directed towards this task." By the defining property of a weakest possible midsequent, if it is not derivable, no midsequent is derivable, and neither is the given sequent $\Gamma \rightarrow \Theta$. The application that Ketonen made of his method of proofs of underivability was to plane projective and affine geometry. The former had been studied in Skolem (1920), a paper universally known for the Löwenheim-Skolem theorem contained in its first part. The second part, mostly forgotten until the 1990s, contains a polynomial-time algorithm for deciding derivability of an atomic formula from a finite number of atomic formulas by the axioms of lattice theory. The third part contains a similar result for plane projective geometry, with the derivability relation between atomic formulas generalized into a finite number of possible atomic cases, a generalization analogous to that in Gentzen's sequent

calculus. Skolem needed it because one of his axioms of projective geometry contains a disjunction in a positive part. To be able to analyze derivations of atomic cases from atomic assumptions purely combinatorially, without the interference of logical operations, he needed the generalization. Skolem's method was to convert the axioms into rules for the construction of formal derivations. The latter do not appear anywhere in print in Skolem's paper, but his verbal explanations amount to as much (p. 116): "The axioms are in fact principles of demonstration." The chapter on Skolem's logic of this *Handbook* contains more details about Skolem's paper.

Ketonen took up in the third part of his thesis Skolem's geometrical work. He used his classical sequent calculus, with the geometrical axioms appearing through their instances in the antecedent parts of sequents. The cut elimination theorem applies in this situation, and Ketonen first repeated Skolem's syntactic proof of the independence of what is known as Desargues' theorem in projective geometry. He then went on to study affine geometry, and gave a similar proof of independence of Euclid's fifth postulate. Skolem's and Ketonen's axiomatizations corresponded both to a purely universal axiom system, which leads to one essential miss in geometry: the axiom of noncollinearity that states the existence of at least three noncollinear points.

How much of Ketonen's work stems from suggestions on the part of Gentzen remains unknown, because no correspondence has been found. Ketonen writes in the preface of his thesis that "Dr. G. Gentzen of Göttingen directed me towards the problem area of this dissertation during my period of study there in 1938–1939." The thesis was Ketonen's only original work in logic, saved from oblivion by a long review that Bernays wrote of it for *The Journal of Symbolic Logic* in 1945. Two somewhat earlier papers in Finnish mention results contained in the thesis, one of them about the completeness of the predicate calculus, the other about proof search with the invertible propositional rules (see Negri and von Plato 2001, pp. 60 and 86 for these, and also von Plato 2004a for more details on Ketonen).

In this account of early uses of Gentzen's sequent calculus, we note Karl Popper's paper (1947). Its review in *Mathematical Reviews* by Evert Beth in 1947 attributes aspects of Popper's calculus to Gentzen, Jaskowski, and Ketonen, even if Popper does not refer to them explicitly. The review shows that Beth knew the calculi of Gentzen as well as the modifications made by Ketonen. When Beth later, in his (1955), presented the tableau calculus, he did not refer to Ketonen but to Kleene (1952). The early work of Hertz, Gentzen, and Popper is discussed at length in Bernays (1965). Invertibility of logical rules is ascribed to Popper, instead of Ketonen, even if Bernays had made the issue completely clear twenty years earlier, in his review (1945) of Ketonen's thesis.

Ketonen's invertible rules were taken to use by Kleene (1952a) in which it is stated that the author knows Ketonen's work only through the Bernays review. Kleene further used them in his influential book of 1952 that also treated intuitionistic sequent calculus in which not all rules can have invertibility. We may consider that with Kleene's book, Gentzen's sequent calculi became generally known and

accessible.

Kleene's work of the early 1950s also pioneered a remarkable development in sequent calculus, namely the "contraction-free" classical and intuitionistic calculi today denoted by *G3c* and *G3i*. These calculi have the property that no structural rules are needed. Weakening is made an admissible rule by letting initial sequents have the form $A, \Gamma \to \Delta, A$, instead of Gentzen's $A \to A$. Contraction is likewise made admissible by a suitable formulation of the rules. If we look at the invertible rules of classical propositional logic, such as in table **22** above, we see how the conclusion uniquely determines the premisses, once the rule applied is specified. Thus, because of the invertibility of the logical rules, "the making of derivations consists of purely mechanical decomposition." These are the words of Ketonen from a paper (1943) in Finnish that first presented the calculus. He must have taken it for granted that the rules of weakening and contraction are not needed in his method of proof search for propositional logic. A calculus for intuitionistic propositional logic with the same kind of termination of root-first proof search was found as late as 1992 by Roy Dyckhoff, and at the same time also by Jörg Hudelmaier.

The classical calculus has the property, mentioned above, of height-preserving invertibility of its logical rules. Albert Dragalin (1988, Russian original 1979) refined the calculus into one in which the structural rules are moreover "height-preserving admissible," meaning that whenever the premiss of such a rule is derivable, the conclusion is derivable without the rule and with at most the same size of derivation. This property has profound effects on cut elimination: In permuting cut up, Gentzen had to restore the original contexts through weakenings and contractions. With the height-preserving admissibility of these rules, the size of a derivation does not increase when the rules are applied. Dragalin (1988) gave also an intuitionistic multisuccedent calculus with the same type of admissibility of the structural rules. Troelsta, finally, gave in Troelsta and Schwichtenberg (2000, first ed. 1996) a single-succedent intuitionistic calculus with height-preserving admissibility of weakening and contraction. The contraction-free sequent calculi are powerful tools for the analysis of formal derivations. Many difficult research results in logic, such as Harrop's theorem, become exercises through the control over the structure of proofs that the *G3*-calculi permit.

The earliest application of sequent calculus in mathematics was in the proof theory of arithmetic, in Gentzen's thesis and in a decisive way in the 1938 proof of the consistency of arithmetic, the latter to be discussed in Section 13. Troelstra mentions Ketonen's work as "an early analysis of cutfree proofs in Gentzen calculi with axioms; but he considers the form of cutfree derivations in the pure calculus where axioms are present in the antecedent of the sequents derived" (Troelstra and Schwichtenberg 2000, p. 142). Another way to apply sequent calculus is to let topsequents in derivations have the form $\to A$, in which A is an axiom, or an instance of a universal axiom. Now, by Gentzen's "extended Hauptsatz," cuts in derivations can be permuted up until one of their premisses is an axiom, but these cuts on axioms remain. Thus, it was believed that "cut elimination

fails in the presence of axioms," as Girard (1987, p. 125) put it. Nevertheless, a method was found by which axioms are converted into rules of sequent calculus of a suitable form. These extend the logical rules of sequent calculus while maintaining the eliminability of cuts. The method was discovered first in the study of some intuitionistic axiom systems for apartness and order, in Negri (1999). It was then shown that the method applies to any universal axiom systems (Negri and von Plato 1998). Finally, Negri (2003) showed that it extends to axiom systems the axioms of which are "geometric implications," a terminology drawn from category theory. This vast class of axioms comprises the usual forms of existential axioms in mathematics.

Extension of sequent calculus by the said method maintains cut elimination, but atomic formulas typically get removed in the application of the rules so that the subformula property of cut-free derivations is lost. Instead, an inessential modification of the subformula property often applies, in which the number of atomic formulas that can get removed is bounded. If we use a classical sequent calculus, such as *G3c*, logical rules can be permuted below the mathematical ones, so that the latter can be considered in isolation. All active formulas are atomic in the rules, and the derivability problem of a mathematical theory concerns in the first place sequents $\Gamma \to \Delta$ in which all formulas are atomic.

Consider a free-variable axiom A, given in conjunctive normal form. Each conjunct can be given as an implication

$$P_1 \& \ldots \& P_m \supset Q_1 \vee \ldots \vee Q_n$$

with $Q_1 \vee \ldots \vee Q_n$ and $\neg(P_1 \& \ldots \& P_m)$ as limiting cases when $m = 0$ and $n = 0$, respectively. These are turned into sequent calculus rules, with two possibilities:

$$\frac{\Gamma \to \Delta, P_1 \quad \ldots \quad \Gamma \to \Delta, P_m}{\Gamma \to \Delta, Q_1, \ldots, Q_n}\, R\text{-}rule \qquad \frac{Q_1, \Gamma \to \Delta \quad \ldots \quad Q_n, \Gamma \to \Delta}{P_1, \ldots, P_m, \Gamma \to \Delta}\, L\text{-}rule$$

Table 23. Axioms converted into sequent calculus rules.

The first type of rule says that if each of the P_i follows as a case from some assumptions Γ, then the Q_j follow as cases, and the second is a dual to this. The latter type of rules, the ones that act on the left part of sequents, are best seen in a root-first order. Assume given a system of such left rules and assume that $\Gamma \to \Delta$ contains only atomic formulas. Now try matching $\Gamma \to \Delta$ as a conclusion to the rules. Whenever there is a match, we have premisses that each get one new atomic formula in the antecedent. Thus, in the end, we obtain the deductive closure of Γ relative to the rules, with branchings into several possible closures each time there is more than one premiss. The succedent Δ remains untouched by the rules. The importance of height-preserving admissibility of contraction is that each rule has to give a new atomic formula in a premiss. In the contrary case, the conclusion would follow by the structural rules from a premiss that has a duplicated formula.

Many theories permit to restrict proof search for a sequent $\Gamma \to \Delta$ with just atomic formulas so that all terms in the derivation are terms known from Γ, Δ.

Derivations with this property are said to have the *subterm* property, in analogy to Gentzen's subformula property of cut-free derivations. There is a bounded number of distinct atomic formulas with known terms, and instances of rules that produce duplications can be excluded by the height-preserving admissibility of contraction. Therefore the search space for a derivation is bounded and the derivability problem decidable. Theories with the subterm property include lattice theory, both in a relational formulation as in Skolem's early work (1920), as well as in a formulation with the lattice operations (Negri and von Plato 2004). In a relational formulation, there are axioms that postulate the existence of the lattice meet (resp. join) of any two elements a, b of a lattice. The subterm property corresponds to the conservativity of these existential axioms over the other lattice axioms, for derivations with the mathematical rules. This conservativity was Skolem's main lemma in his early proof of the decidability of derivability with the axioms of lattice theory. In a formulation with operations, we have for any two elements a, b a meet $a \wedge b$ and a join $a \vee b$. The subterm property means that no new objects have to be constructed, by which decidability of derivability with the lattice rules follows. We note that first-order lattice theory is not decidable.

Skolem (1920) established a result for plane projective geometry analogous to that for lattices. As mentioned, this result was redone within sequent calculus by Ketonen (1944) who also extended it to affine geometry, but neither was able to include the axiom of noncollinearity in their systems. The subterm property for plane projective and affine geometry with noncollinearity included, treated by the method of extending sequent calculus with mathematical rules, was established in von Plato (2004). The subterm property for these geometries has the effect that the existence of at least three non-collinear points is conservative over the rest of the geometrical rules: A proper use of an existential axiom requires an existential theorem, one could say. Decidability of derivability by the geometrical rules follows, but as with lattice theory, the first-order theory remains undecidable.

11 GENTZEN'S FIRST PROOF OF THE CONSISTENCY OF ARITHMETIC

Gentzen's thesis presented his analysis of the structure of derivations in pure logic. The work was finished by the spring of 1933, and his next task, or actually the task he had set aside to write up his doctoral thesis, was to find a proof of the consistency of elementary arithmetic through an analysis of arithmetic derivations. This he succeeded in doing by the end of 1934, and the result was submitted for publication in the following summer in a long paper with the title *Die Widerspruchsfreiheit der reinen Zahlentheorie* (The consistency of pure number theory).

Criticisms by Bernays and Gödel made Gentzen change his original proof into one that uses the now famous method of transfinite induction up to the ordinal ϵ_0. As is indicated in the published paper of 1936, the most crucial part of the proof was changed early in 1936. The original proof of consistency of arithmetic that was submitted to *Mathematische Annalen*, the leading journal of mathematics at

that time, was preserved in the form of galley proofs. An English translation of the parts differing from the 1936 version was published in Szabo's edition of Gentzen's papers in 1969. The German original was published in 1974 with the title *Der erste Widerpruchsfreiheitsbeweis für die klassische Zahlentheorie* (The first proof of consistency for classical number theory) and with a short introduction by Bernays.

As explained in Bernays' paper *On the original Gentzen consistency proof for number theory,* of 1970, the criticism was that the proof "implicitly included an application of the fan theorem." In 1970, he held his own earlier criticisms doubtful. Kreisel (1987, p. 174), in fact, reports a result by which the addition of the fan theorem to Peano arithmetic would not make its consistency provable. The proof is based, instead, on an induction over well-founded infinitely branching trees, or the "bar theorem" in other words.

Gentzen's original proof of 1935 is the easiest one to understand of his altogether four different proofs of the consistency of arithmetic. It uses a system of natural deduction for classical logic, and the basic idea can be described as follows: Assume that a false formula C is derivable from some assumptions Γ. Think of an attempt at falsifying C through a "worst possible" situation. Say, if C is $\forall x A$, a falsifying instance $A(n/x)$ is chosen. The task is to show that in this situation, also some formula in Γ can be falsified. Then, if C is derivable from Γ and no formula in Γ can be falsified, also C is unfalsifiable.

Gentzen's 1935 system has just the rules for the connectives $\neg$ and $\&$ and the universal quantifier, written in sequent calculus style. The rules for $\&$ and $\forall$ are as in tables **11** and **12**. The rules for negation are:

$$\frac{\Gamma \to \neg\neg A}{\Gamma \to A}\,DN \qquad \frac{A, \Gamma \to B \quad A, \Delta \to \neg B}{\Gamma, \Delta \to \neg A}\,Wid$$

Table 24. Gentzen's 1935 negation rules.

Wid stands for the German "Widerlegung," or refutation of the assumption A that leads to contradictory conlusions. The rule is substantially the same as a rule of negation introduction. Gentzen notes (12.4) that the rule of double negation could be left out, because it has the same effect as the rule of indirect proof that is admissible.

The induction rule is

$$\frac{A(y/x), \Gamma \to A(sy/x)}{A(0/x), \Gamma \to A(t/x)}\,Ind$$

Table 25. Gentzen's induction rule in sequent calculus style.

Here y is an eigenvariable, not free in the conclusion of the rule, and t is any term. The rest of the arithmetical axioms are treated as in Section 13 below, and here we consider only the logical rules and the induction rule. Gentzen does not deal with the question of the closure with respect to composition of derivations in his system of rules.

The attempt at falsification can be described, let us hope not too anachronistically because the idea is so immediate, as a game against an omniscient "falsifier," and the consistency of arithmetic follows from the availability of a "winning strategy" in such a game. Given a sequent $\Gamma \longrightarrow C$, a *reduction procedure*, in Gentzen's words, is defined for it. Reduction is effected by suitable moves: First certain "S-moves" that operate on the succedent part are taken by the falsifier. One tries to respond to these by certain "A-moves" on the antecedent part of the given sequent. The former moves are determined by the form of the given sequent, as follows:

SVar. $\Gamma \longrightarrow C$ *has free variables. These are instantiated at will.*

S$\&$**.** $\Gamma \longrightarrow A\&B$. *One of* $\Gamma \longrightarrow A$ *or* $\Gamma \longrightarrow B$ *is chosen.*

S$\neg$**.** $\Gamma \longrightarrow \neg A$. *The reduced sequent is* $A, \Gamma \longrightarrow 0 = 1$.

S$\forall$**.** $\Gamma \longrightarrow \forall x A$. *Some instance* $\Gamma \longrightarrow A(k/x)$ *is chosen.*

Table 26. Gentzen's reduction rules in the succedent part.

The *order of precedence* among the **S**-moves is that move **SVar** comes before the other **S**-moves. The falsifier knows how to end up with the worst possible case, here, a false equation as a conclusion. Each step the falsifier takes simplifies the succedent of the sequent to be reduced until an equation $m = n$ remains. If the equation is true, the falsifier lost. Otherwise, when no **S**-move is applicable and $m = n$ is false, one tries to show that some of the assumptions must contain a falsity, too. To do this, the following moves are available:

A$\&$**.** *The sequent is* $A\&B, \Gamma \longrightarrow m = n$ *with* $m = n$ *false. For the reduced sequent, one of* $A, \Gamma \longrightarrow m = n$ *and* $B, \Gamma \longrightarrow m = n$ *is chosen.*

A$\neg$**.** *The sequent is* $\neg A, \Gamma \longrightarrow m = n$ *with* $m = n$ *false. The reducent sequent is* $\neg A, \Gamma \longrightarrow A$.

A$\forall$**.** *The sequent is* $\forall x A, \Gamma \longrightarrow m = n$ *with* $m = n$ *false. For the reduced sequent* $A(k/x), \forall x A, \Gamma \longrightarrow m = n$ *with some* k *is chosen.*

Table 27. Gentzen's reduction rules in the antecedent part.

The order of precedence is that **S**-moves come always before **A**-moves. The aim is to produce a false formula in the antecedent part of a sequent, whenever the falsifier has produced one in the succedent. Note that if a negation at left is reduced, the falsifier is in turn, unless it was a negation of an equality. Given a sequent $\Gamma \longrightarrow C$, the result of the falsification moves is a sequent to which no such move applies. Thus, the aim is to construct a winning strategy in the falsification game, to meet whatever choices the falsifier may make.

A sequent is defined as *irreducible* if no reduction move applies to it. A sequent $\Gamma \longrightarrow m = n$ is in *endform* if either $m = n$ is true or there is some false equality in Γ. Finally, a sequent $\Gamma \longrightarrow C$ is *unfalsifiable* if for any choices of **S**-moves, there are **A**-moves such that $\Gamma \longrightarrow C$ reduces to endform.

Gentzen's consistency proof shows that all derivable sequents are unfalsifiable. It follows that the sequent $\longrightarrow 0 = 1$ is not derivable, because it is irreducible but not in endform: No equation in the antecedent is false, because there are none.

The proof proceeds by first showing, by induction on formula length, that initial sequents reduce to endform. Next a crucial lemma shows that if two sequents $\Gamma \longrightarrow A$ and $A, \Delta \longrightarrow C$ are reducible to endform and possible eigenvariables distinct, their *composition* into $\Gamma, \Delta \longrightarrow C$ is reducible to endform. The proof is by induction on the length of the composition formula A. Next it is shown by induction on the height of a derivation that if the premisses of the logical rules and the induction rule reduce to endform, also their conclusions reduce to endform. It then follows that if the sequent $\Gamma \longrightarrow C$ is derivable, it is unfalsifiable. Consistency is an immediate corollary: The sequent $\longrightarrow 0 = 1$ is irreducible but not in endform, and therefore it is not derivable.

Our description of Gentzen's proof is straightforward enough. However, Gentzen does not always state in any clear terms how his proofs are structured, or even what precisely is being proved. Such things have to be worked out by the reader. There are other kinds of loose ends, as well: As mentioned, closure with respect to composition of derivations in Gentzen's rule system is not treated. Secondly, Gentzen does not use any normal form in his proofs, which makes them complicated. Lack of such a form is reflected in the use of the crucial lemma, closure of reduction to endform under composition, in the proof of the main result about the reducibility to endform of derivable sequents. Composition has to be used not only with the rule of induction, contrary to what one would expect, but also with the logical rules *DN* and *Wid*, which can be avoided through the use of a normal form for derivations. (For a detailed proof along the lines sketched here, see von Plato 2005.)

The reduction procedure gives an infinity of choices when $\mathbf{S}\forall$ is applied, thus a winning strategy has to be defined inductively in an infinitely branching tree.

12 CRITICISMS OF GENTZEN'S CONSISTENCY PROOF

Gentzen's original consistency proof was the most important result in logic and foundations after Gödel's incompleteness theorems. It aroused immediate interest, especially through Bernays who stayed at the Institute for Advanced Study in Princeton in 1935–36. He had sailed in September 1935 over the Atlantic, in the company of Gödel and with Gentzen's manuscript in his briefcase. Correspondence between Bernays and Gentzen documents to some extent the objections to Gentzen's first proof. Only Gentzen's letters have been preserved. Gödel made a suggestion to change the proof, of which we only know, in Gentzen's words, that "the modification suggested by Gödel was known to me, but it is in reality not usable from the finitary point of view, due to its impredicative character." (Menzler-Trott 2001, p. 58, and also Kreisel 1987, pp. 173–174). At that time, Gentzen took his proof to be finitary, even if not representable in arithmetic. He writes in the original version that the methods used in the proof are undoubtable

("unbedenklich") or "at least essentially more secure than the doubtable forms of inference of pure number theory." The reason is that his methods are "finitary in a specific sense," as defined in his paper. "That this is the case shall be considered more closely in Section V." (p. 98 of the German version). In the printed paper, the last passage was changed into: "To what extent this is the case..."

After the criticism of Bernays, Gödel, and possibly also von Neumann, Gentzen changed his original proof. The reason was not that there was objectively something to criticize: As mentioned, it can be shown that the "fan theorem" Bernays thought to be an illicit moment behind Gentzen's proof, would not make the consistency of Peano arithmetic provable. Before any criticisms and before the first proof, Gentzen had conceived a different argument for the termination of the reduction procedure, one that uses transfinite induction. A specific notation is created to this purpose in the published version of 1936; However, in his two later proofs, of 1938 and 1943, Gentzen used the standard notation for ordinal numbers. Gentzen's ordinal ϵ_0 can be defined as follows in such a notation: Let ω stand for the ordinal of the natural numbers, and let $\omega_0 \equiv 0$ and $\omega_{n+1} \equiv \omega^{\omega_n}$. The ordinal ϵ_0 is the limit of ω_n when n grows to infinity.

The 1936 paper suffered modifications in two long sections, which created incongruities with earlier parts that remained unchanged. At the same time, Gentzen took the occasion to simplify the proof through a modification of the system of natural deduction, by replacing most of the logical rules with "axioms" as in the *Ms.ULS* (see Section 8 above). Last, the general procedure for the reduction of sequents was not used at all, but a new one for the reduction of derivations.

In the modified second proof of the consistency of arithmetic, the number of logical rules of natural deduction is cut down by admitting the following sequents among the initial sequents:

$$A\&B \rightarrow A \quad A\&B \rightarrow B \quad A, B \rightarrow A\&B \quad \forall x A \rightarrow A(t/x)$$

$$A, \neg A \rightarrow 0 = 1 \quad \neg\neg A \rightarrow A$$

Table 28. The initial sequents of Gentzen's 1936 calculus.

Of the logical rules only $\forall I$ and $\neg I$ (Gentzen's *Wid*) remain. All the other ones follow from the added initial sequents through Gentzen's "chain rule":

$$\frac{\Gamma_1 \rightarrow A_1 \quad \ldots \quad \Gamma_n \rightarrow A_n \quad A_1, \ldots A_n, \Delta \rightarrow C}{\Gamma_1, \ldots \Gamma_n, \Delta \rightarrow C}\, Chn$$

Table 29. Gentzen's chain rule.

In terms of standard natural deduction, the effect of the chain rule is that the open assumptions $A_1, \ldots A_n$ of a derivation are substituted through their derivations in a simultaneous substitution.

In Bernays (1970), a proof of consistency is sketched that uses the idea of Gentzen's unpublished proof, but the logical calculus is not the one of the original proof. It is, instead, the above modification taken from the printed proof

of 1936. The reader has some difficulty in gathering these details from Bernays' presentation.

Bernays wrote a review (1936a) of the published proof that explains in a factual way the central method of the proof. He then states that by Gödel's incompleteness theorem, the transfinite induction principle cannot be provable in arithmetic, whereas inductions to strictly below ϵ_0 would. How the latter follows is not indicated.

Gödel's struggles with Gentzen's proof became public only in 1958, with the appearance of his paper *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes* (On a so far unused extension of the finitary position) in the journal *Dialectica*. The third volume of Gödel's *Collected Works*, with the previously unpublished lectures and papers, shed a lot more light on his involvement with the consistency of arithmetic. Gentzen's proof must have come as a surprise to him. Nothing in his own work indicates that he would have been preoccupied with a constructive proof of consistency. By the end of 1938, instead, he had distinguished three ways to treat the consistency problem. These are found in the "Zilsel" lecture in the third volume: the use of higher-order primitive recursive functionals, the use of provability logic, and the use of transfinite induction. The "Yale" lecture of 1941, in the same volume, is a finished presentation of Gödel's position and his consistency proof for arithmetic. As its title *In what sense is intuitionistic logic constructive?* indicates, Gödel had doubts about the intuitionistic explanation of logic. Especially, the explanation of a proof of an implication was considered circular, as discussed towards the end of Section 6 above. Thus, Gödel formulated a "logic-free" hierarcy of primitive recursive arithmetic functionals that can interpret the intuitionistic logic used in Heyting arithmetic. Consistency of arithmetic reduces to the computability of terms in Gödel's functional calculus.

One may ask why Gödel did not publish his own alternative proof of the consistency of arithmetic at the time, in the early forties. Gödel's relative silence is in contrast with the picture one gets from his *Arbeitshefte*: As seen from Dawson and Dawson (2005), these are full of notes on work concerned with Gentzen's proof, treated from different points of view. One of Gödel's reasons for not publishing may have been the "shifting interests towards philosophy" that he later recollected in connection with the work on the independence of the axiom of choice from the same time. However, the interesting question from a systematic point of view is what real novelty there is in Gödel's proof. Isn't it clear that if the computability of terms in a hierarchy of primitive recursive functionals secures the consistency of arithmetic, then there should be somewhere hidden a principle equivalent to, or even stronger than, Gentzen's transfinite induction to the ordinal ϵ_0? That the two methods are equivalent from this point of view follows from Tait's (1967) interpretation of Gödel's theory of primitive recursive functionals in Heyting arithmetic. Tait (2001,2005) offers profound discussions and comparisons of Gentzen's proof and Gödel's 1938 view of it.

The deepest foundational issue in the 1930s was the incompleteness of mathematics and its aftermath, in which latter Gödel and Gentzen were the *dramatis*

personae. Gentzen never learned about Gödel's way out, unless the latter described it when they came to know each other in December 1939. Still, Gentzen had the last word anyway, by his last paper of 1943. The closest Gödel came to an appreciation of Gentzen's determination of the proof-theoretical strength of Peano arithmetic, doubtless the most important step ahead after his own paper of 1931, seems to be a passage in a letter of March 1966 to Constance Reid who at the time was preparing her Hilbert-biography. He first states (*Collected Works*, vol. V) that Hilbert had tried to prove the consistency of classical mathematics "on the basis of concrete evidence" and continues that "viewing the situation from a purely *mathematical* point of view, consistency proofs on the basis of suitably chosen stronger metamathematical presuppositions (as they have been given by Gentzen and others) are just as interesting, and they lead to highly important insights into the proof theoretic structure of mathematics."

13 THE NEW PROOF OF CONSISTENCY OF 1938

The main novelty of Gentzen's second proof of the consistency of arithmetic was the assignment of ordinal numbers to sequents in derivations and the proof that each reduction step on derivations diminishes the ordinal of the endsequent. However, Gentzen was not satisfied with the way the proof came to be presented. His style, as such, creates difficulties with its long, verbose arguments the details of which have to be laboriously worked out by the reader. Added to this come the incongruities created by substantial changes that were not worked through for the whole paper. Thus, in 1938, Gentzen returned to the topic and published a third proof of the consistency of arithmetic. Its aim was, in his own words, "to detail out the basic ideas and to make each single step of the proof as understandable as possible." In a letter to Bernays of 12 May 1938, Gentzen writes about his paper that "it should be fairly clearly visible from the new presentation how I arrived at the consistency proof, through the method of proof of my dissertation." (from Menzler-Trott 2001, p. 79).

The logical calculus is the classical sequent calculus *LK* of Gentzen's thesis, with implication left out. The rule of induction is as in table **25** above, but with an arbitrary context Θ added in the succedent. The rest of the arithmetical axioms are converted into "mathematical basic sequents." Each of these axioms is of the form $P_1 \& \ldots \& P_n \supset Q$ in which the P_i and Q are equations, with possibly $n = 0$. These implications are turned into sequents of the form $P_1, \ldots, P_n \rightarrow Q$. The "extended *Hauptsatz*" of Gentzen's thesis states that if among the topsequents of derivations mathematical basic sequents are permitted, the rule of cut can be permuted up so that it is applied only to formulas in such sequents. Then the question of the consistency of arithmetic concerns just the logical rules (together with the structural rules) and the rule of induction. The rule of cut is indispensable among the structural rules.

Gentzen's "new proof" of 1938 shows directly that there cannot be any derivation of the empty sequent $\rightarrow$. The idea of the proof is to show that if there is

such a derivation, it can be simplified. First a procedure is given for the conversion of any derivation of the empty sequent into a simpler form. Next it is shown that the conversion steps terminate with something that cannot be a derivation of the empty sequent. The termination proof is based on the assignment of ordinals to sequents occurring in a derivation, and to a proof that each conversion step decreases the ordinal of an empty endsequent. In the end, a derivation of $\rightarrow$ exists only if $\rightarrow$ is a conclusion of a logical rule, which it is not.

The proof analyzes the *endpiece* (*Endstück*) of a derivation, defined as the subderivation obtained by starting from the endsequent and including each sequent until an initial sequent or a conclusion of a logical rule is reached. The proof begins with the following: If there is an instance of rule *Ind* in the endpiece and if the term t in its conclusion is a variable, it can be substituted by 0. Otherwise it is some number i and we have:

$$\frac{A(y/x), \Gamma \rightarrow \Theta, A(sy/x)}{A(0/x), \Gamma \rightarrow \Theta, A(i/x)}\, Ind$$

The sequent $A(y/x), \Gamma \rightarrow \Theta, A(sy/x)$ is derivable for any numerical value of y. The induction up to i is replaced by $i-1$ cuts on such instances, as in

$$\dfrac{\dfrac{A(0/x), \Gamma \rightarrow \Theta, A(1/x) \quad A(1/x), \Gamma \rightarrow \Theta, A(2/x)}{A(0/x), \Gamma^2 \rightarrow \Theta^2, A(2/x)}\, Cut \\ \vdots \\ \dfrac{A(0/x), \Gamma^{i-1} \rightarrow \Theta^{i-1}, A(i-1/x) \qquad A(i-1/x), \Gamma \rightarrow \Theta, A(i/x)}{A(0/x), \Gamma^i \rightarrow \Theta^i, A(i/x)}\, Cut$$

Here the exponents denote equally many copies of the contexts. The copies are eliminated by contractions until the conclusion of the original *Ind* is recovered. After inductions up to numbers have been replaced by cuts, the endpiece contains only structural rules.

A crucial lemma (in paragraph 3.4.3 of Gentzen's paper) shows that there are at least two sequents beginning the endpiece such that a cut formula of the endpiece is principal in a left and a right rule, respectively. The purported proof of this lemma is a long, verbose argument from which it is very difficult to see how it actually establishes what the lemma claims. An alternative proof is contained in Takeuti's book *Proof Theory* of 1975.

The difficulty in Gentzen's lemma stems from the presence of other structural rules than cut in the endpiece. If the endpiece has only cuts, with n distinct cut formulas, there are $n+1$ leaves in the endpiece so that either one cut formula appears at least twice on the same side of at least one sequent, and could be contracted and the number of cuts reduced, or at least one cut formula is principal on a left and right side, respectively, of a leaf in the endpiece.

By the lemma, at least one cut in the endpiece can be simplified by replacing it with cuts on shorter formulas, just as in our first example of permutation of cut in Section 7. Here we see at work the "method of proof of my dissertation," as

Gentzen wrote in the letter of May 1938 to Bernays quoted above. However, the matter is not as simple in 1938 as it was in the doctoral thesis: After the cut has been permuted up into cuts with shorter formulas, cuts with these formulas have to be repeated lower down in the derivation. It is crucial that in these cuts, the cut formula be principal only in the right premiss. Now the concepts of *Höhe* and *Höhenlinie* are introduced: The *height* of a sequent in a derivation is the maximal length of a cut or induction formula below the given sequent in the derivation. The empty endsequent has height 0, and *height-lines* indicate places at which the height gets diminished. The effect of Gentzen's overall transformation (1938, 3.5) is that the height-line gets, so to say, pulled up in the derivation. We shall not go into the complicated details of this proof, save to say that still today, no proof without the *Höhenlinie* and with the same ordinal ϵ_0 as in Gentzen's proof seems to have emerged. (We owe this information to Wilfried Buchholz.)

The formal proof of termination of the reduction process for derivations of the empty sequent uses an assignment of ordinals to sequents in derivations, and the verification that each reduction step decreases the ordinal of the empty endsequent.

Gentzen's third proof of the consistency of arithmetic gave a concise and convincing presentation of the result. Others offered soon alternatives to Gentzen's treatment of the matter in terms of sequent calculus: Laszlo Kalmar gave in 1938 a proof of consistency within the axiomatic tradition of logic, explained in the second edition of the second volume of Hilbert and Bernays' *Grundlagen der Mathematik* (1970). The original paper can be found in the Bernays archives at ETH-Zurich. In 1940, Wilhelm Ackermann published a proof based on some ideas of Hilbert's, ideas that Ackermann himself had applied in his proof of consistency of arithmetic without the principle of induction in the mid-1920s. In the late 1930s, after solicitations on the part of Bernays, Ackermann recast Gentzen's proof in terms of Hilbert's so-called ϵ-calculus (the name is a coincidence and has nothing to do with the ϵ of Gentzen's induction principle).

14 GENTZEN'S LAST PAPER

Gentzen finished in 1939 what is known as a *Habilitationsschrift*, a kind of second thesis by which one establishes oneself as part of a university department in Germany. The script became Gentzen's last paper published during his lifetimes: the *Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie.* (Provability and unprovability of initial segments of transfinite induction in pure number theory). It came out in the *Mathematische Annalen* in 1943, in a last issue to appear for years. Quite amazingly, a review of the paper appeared in the United States during the war, in 1944. The reviewer was Bernays who had moved to Switzerland after 1933. He worked as a bridge between European and American logicians through the late thirties and the war: The present author has seen a letter from Bernays to Georg Henrik von Wright who wrote reviews for *The Journal of Symbolic Logic* in Helsinki during the war. Bernays asks for second copies of the reviews, because the air mail letters with the

original reviews "were destroyed in an air attack."

Despite Bernays' efforts, the methods of Gentzen's 1943 paper, or the earlier 1938 paper, did not become widely used: Both works remained by and large unread in the 1940s and early 1950s, except for a chosen few such as Schütte, Kreisel, and Takeuti. Kleene's *Introduction to Metamathematics* had covered the logical aspects of sequent calculi, but no accessible introduction to the proof theory of arithmetic was available before Takeuti's *Proof Theory* of 1975. Even today, one can find whole books devoted to Gödel's incompleteness theorem, both popular and scientific, that do not as much as mention Gentzen, the person who has contributed most to solving the problems of consistency and incompleteness of arithmetic.

Gentzen's idea in his last paper was to extend the domain of natural numbers by constructive ordinals and to formulate transfinite induction as a separate rule of proof within the extended system. The basis was, as in 1938, his classical sequent calculus *LK* of the doctoral thesis of 1933. He then proceeded to prove that transfinite induction up to ordinals below ϵ_0 is an admissible rule of the sequent calculus for arithmetic with ordinary induction. This observation, apparently due in some form to Bernays, was made already in 1936.

As a second and main result, Gentzen showed that transfinite induction up to ϵ_0 is not provable by ordinary arithmetic, even if it is expressible in arithmetic. By the unprovability of transfinite induction, arithmetic is consistent, because anything would be provable in an inconsistent system. Secondly, arithmetic is incomplete: Transfinite induction expressed as an arithmetic formula is a true but unprovable sentence. As Gentzen points out, transfinite induction is a mathematical principle, whereas Gödel's original proof of incompleteness had produced a "metamathematical" principle that was coded into arithmetic through Gödel numbering. Thus, it was often held to be an artificial construction that has no serious bearing on "real mathematics."

The type $\mathcal{O}$ of *ordinal numbers below* ϵ_0 is defined by the following inductive clauses. 1. $0 : \mathcal{O}$. 2. If $\alpha : \mathcal{O}$ and $\beta : \mathcal{O}$, then $\alpha + \beta : \mathcal{O}$. 3. If $\alpha : \mathcal{O}$, then $\omega^\alpha : \mathcal{O}$. The basic relations among ordinals are $\alpha = \beta$, $\alpha \prec \beta$, and $\alpha \preceq \beta$. The following notation is used: $\omega_0 \equiv 0, \omega_1 \equiv 1 \equiv \omega^0, \omega_2 \equiv \omega, \ldots, \omega_{\alpha+1} \equiv \omega^{\omega_\alpha}$ if $\alpha \prec \omega$, and $\omega_\alpha \equiv 0$ if $\alpha \not\prec \omega$.

The place of natural numbers is taken by the *finite ordinals*: those α for which $\alpha \prec \omega$. The successor function for natural numbers is represented as $sn \equiv n+\omega^0 \equiv n+1$. Ordinals for which $\alpha \prec \omega$ is not the case are *transfinite ordinals*. The rule of induction is as follows:

$$\frac{y \prec \omega, A(y/x), \Gamma \longrightarrow \Theta, A(sy/x)}{t \prec \omega, A(0/x), \Gamma \longrightarrow \Theta, A(t/x)}\, Ind$$

Table 30. Gentzen's 1943 rule of induction.

As in the rule of table **25**, y is the eigenvariable. The only difference with the earlier rule is that the finiteness conditions $y \prec \omega$ and $t \prec \omega$ have been added in the assumption parts.

The rule of *transfinite induction* is

$$\frac{y \succ 0, \forall z(z \prec y \supset A(z/x)), \Gamma \rightarrow \Theta, A(y/x)}{A(0/x), \Gamma \rightarrow \Theta, A(t/x)}\ TI$$

Table 31. Gentzen's rule of transfinite induction.

Also here, y is the eigenvariable of the rule. If t is a numerical term (an ordinal less than ϵ_0), application of *TI* gives a "derivation of transfinite induction up to the ordinal t."

Gentzen shows first that rule *TI* is not needed below ϵ_0. To this purpose, he introduces a one-place "property parameter" $\mathcal{P}$. When $\mathcal{P}(0) \rightarrow \mathcal{P}(t)$ is the endsequent of a derivation, we have "derivability of transfinite induction up to t." Gentzen proves that if t is a numerical term, then $\mathcal{P}(0) \rightarrow \mathcal{P}(t)$ is derivable from the premiss $t \succ 0, \forall y(y \prec t \supset \mathcal{P}(y)) \rightarrow \mathcal{P}(t)$ of rule *TI* in the calculus *LK*+*Ind*.

The main result is that transfinite induction up to ϵ_0 is not derivable in the calculus *LK*+*TI*, unlike inductions to lesser ordinals. The proof is based on the reduction procedure of the third, 1938 proof of consistency of arithmetic. First the notion of an endpiece is extended to *LK*+*TI*. In 1938, the endpiece ended with an empty sequent; now instead, the endsequent has the form $\mathcal{P}(0) \rightarrow \mathcal{P}(t_1), \ldots, \mathcal{P}(t_n)$. The next step is to assign ordinals to derivations of such endsequents. A crucial lemma states that in a derivation of an endsequent, the least of the ordinals t_i cannot be greater than the ordinal of the derivation, or else the derivation can be reduced. The ordinals of derivations are less than ϵ_0, so it follows that transfinite induction up to ϵ_0 is not derivable. A more precise result can be given: In a derivation of $P(0) \rightarrow P(t)$, the maximum length n of induction formulas in the derivation has the following connection to t: $t \prec \omega_{n+3}$.

Gentzen's last paper marked the beginning of the second main direction of proof theory in addition to structural proof theory, namely *ordinal proof theory*. The *proof-theoretical ordinal* of a theory is defined as the least ordinal that is needed for proving its consistency, just as with ϵ_0 for Peano arithmetic. The central aim in ordinal proof theory has been to determine the proof-theoretical ordinals of ever stronger subsystems of second-order arithmetic. Among the highlights of this development is the determination of the "ordinal of predicativity" around 1964, independently of each other, by Schütte and Solomon Feferman. The present state is marked by ordinal analysis of the principle of Π^1_2-comprehension in the work of Toshiyasu Arai and Michael Rathjen. These developments are reviewed in Rathjen (1995) and (1999).

15 MISCELLANEA

We recall here briefly the contents of those papers of Gentzen that have not been discussed above.

Gentzen published in 1936 a proof of the consistency of Russell's simple theory of types. The same result was proved by Alfred Tarski (1933), as Gentzen was

eager to point in a few lines he wanted added to Schmidt's (1936) review of his paper. The introduction to Gentzen's paper describes the consistency problem as a part of Hilbert's program: "Simple type theory comprises essentially the system of *Principia Mathematica* together with the axiom of choice but without the axiom of infinity. It is only through the inclusion of this axiom ... that essential difficulties arise, difficulties the solution of which is the main task of Hilbert's proof theory." Gentzen gives an axiomatic treatment of simple type theory and constructs a model for it starting with just one ground element. Each level in the type structure then has a bounded number of elements and the truth values of derivable formulas can be calculated. The motive for studying type theory should have been that it is one possible framework for the formalization of mathematics beyond arithmetic. Gentzen did not, however, pursue this line further.

The last paper that Gentzen wrote, dated 19 December 1944, is a short piece dedicated to Heinrich Scholz' 60th birthday. It was published posthumously as Gentzen (1954). The paper shows that arithmetic derivations with several applications of the principle of induction can be converted into derivations with just one such application. Gentzen gained from the result the insight that in the analysis of derivations in arithmetic, the number of inductions is not essential, but the length of the longest induction formula, an insight that was important for Gentzen's 1943 paper on transfinite induction.

Gentzen published in 1938 a philosophical paper on "the present state in the research on foundations of mathematics." Two earlier papers contain similar materials, namely Gentzen (1936–37), and the lecture at the philosophy congress in Paris, Gentzen (1937). The leading idea in these presentations is that finitary methods are not sufficient for the purposes of consistency proofs in mathematics, and have to be replaced by stronger, still constructive principles, as found in intuitionistic mathematics. Gentzen thought that a proof of the consistency of analysis, i.e., of the theory of real numbers, could be obtained, but was more sceptical about the possibilities for such a proof for set theory.

Gentzen published a number of reviews that can be found listed, and most of them reproduced, in Mentzler-Trott (2001, p. 311). A review of Emil Post's 1936 formulation of the notion of a recursive function, in the paper *Finite combinatory processes*, mentions other definitions of recursiveness by Church, Gödel and Herbrand, and Kleene. Gentzen then notes that a process is finite if it ends in a finite number of steps. In this connection, he points at the problematic aspect of the notion of recursiveness by putting in italics his final comment: "*Of what kind the proof of finiteness of the process should be is, as also in Church, not specified.*"

The review of Barkley Rosser's 1937 paper on Gödel's theorem discusses briefly attempts at formulating infinitary rules of inference, and mentions earlier uses of infinitary rules in the work of Hilbert and Rudolf Carnap.

16 THE GENTZEN PAPERS

After the war, there were attempts at locating Gentzen's papers, especially a presumed proof or near-proof of the consistency of analysis. The accounts of Gentzen's last days tell that he was thinking constantly of the consistency problem and had made some essential breakthroughs (See Menzler-Trott 2001, p. 281). Only a short stenographic memorandum written by Gentzen in February 1945 was found, transcribed in 1948. Bernays read it at the time but was not able to make any opinion in the matter. The manuscript referred to another thing called "WA," for "Widerspruchsfreiheit der Analysis" (consistency of analysis) of which it was a summary, as it says, "through page 339." One can double or triple that to get an idea of how many printed pages such a set of stenographic notes would correspond to.

A letter of Arnold Schmidt to Gentzen's mother, from 1948, states that there are manuscripts in Göttingen that are preliminary work for Gentzen's published papers. It is suggested that these would not contain anything of new scientific value, and that the papers be placed in an archive together with Hilbert's papers. No such papers have been, alas, found in Göttingen. It is assumed that the papers were destroyed together with Schmidt's papers, the latter at his will after his death in 1964.

It turned out twenty years after the last-mentioned incident that Gentzen's sister Waltraut Student was in the possession of two folders of Gentzen's shorthand notes. These were given to Christian Thiel who, on the basis of the opinion expressed by the editor of Gentzen's *Collected Papers*, thought that the papers would not contain anything that would not have been known already, and would thus be of purely historical interest. On this basis, he has been transcribing the papers occasionally and is close to halfway through early in 2005. The papers will be made available together with the *Ms. ULS* and the February 1945 summary of *WA* once the editing of these materials is complete.

The Gentzen papers were selected out and organized by Gentzen as being of possible later use for him. The last items are from the late summer of 1944. Gentzen may have thought that the papers were in a safer place if not in Prague. He took them to his mother's place on the Baltic island of Rügen.

The Gentzen papers consist of stenographic manuscripts that should amount to well over 500 printed pages. They are divided into a "blue" and a "violet" folder and are written between 1931 and 1944. They are further divided into various series with page numberings and, usually, a date here and there. These manuscripts give a unique insight into the way Gentzen worked. With the birth of natural deduction, for example, he would work through meticulously alternative possibilities, such as the treatment of variable conditions, substitution, discharge of assumptions, and so on. From the extensive notes he would distill work to be published, altogether 11 papers of exceptional quality in respect of novelty and execution (except perhaps the 1936 consistency paper that was modified at proof stage with a result that did not satisfy him). The most essential feature of these

papers is their development of new ideas without any errors being committed.

We give here a chronological description of the different series. Occasional notes from later times are sometimes added in between the pages belonging to a series.

The first series is designated *VOR*. The remaining few pages, with numbers between 13 and 20, have been written in the fall of 1931, so after the work on Hertz systems in the preceding summer. Gentzen treats mainly the question of functions *vs.* relations, i.e., the formulation of axiomatic theories with operations against a formulation with relations. The topic is treated in Hilbert and Ackermann's book of 1928 which was Gentzen's main source early on. A relational axiomatization of (parts of) arithmetic is included in *VOR*. The method is also applied in Gentzen's thesis in the proof of consistency of arithmetic without the principle of induction. Gentzen kept an interest in the topic also later, which is why he kept the notes.

Gentzen's work towards a consistency proof of arithmetic and analysis begins early in 1932. By September, he had formulated his system of natural deduction, with an induction rule included, and had conjectured (if not proved) the "Hügelsatz" we discussed in Section 5. Some pages refer to a lost series "HÜG" with pages like 62.

Another disappeared series had the signum *D* (for *Dissertation*?). From p. 94 on, it becomes the preserved series *INH* that had been kept complete and bound together into a single piece with new page numbers 1–36. The full title is *Die formale Erfassung des Begriffs der inhaltlichen Richtigkeit in der reinen Zahlentheore* (Formal conception of the notion of intuitive-concrete correctness in pure number theory). Two-thirds of *INH* have been written in about a month, starting in October 1932. At this time, Gentzen still believed to obtain a consistency proof for arithmetic soon. His thought seems to have been, also on the basis of *Ms.ULS*, that the normalization theorem and subformula property of natural deduction would extend to derivations of a contradiction in arithmetic. He found out later that one cannot restrict the induction rule into known formulas.

The fall of 1934 sees a continuation of the *INH* series. At the same time, Gentzen began also a series *BZ*, for *Beweistheorie der Zahlentheorie*. This one runs to January 1935 when he already had the proof of the consistency of arithmetic finished.

There are other shorter and fragmentary series that deal with topics from propositional and predicate logic and from simple type theory, especially the role of the axiom of choice in the last mentioned. One series of notes is called *AL* (for *Aussagenlogik*) and has materials from various times. One that has so far been deciphered contains a semantical decision procedure for intuitionistic propositional logic. Gentzen sets out to define validity by the conditions that $A\&B$ be valid if both A and B are valid, $A \vee B$ be valid if one of A and B is valid, and $A \supset B$ be valid if B is valid whenever A is valid. The only difficulty is when there are iterated implications in the antecedents, as in $(A \supset B) \supset C$. Here C has to be shown valid under the condition that B be valid whenever A is. One considers never the case that the antecedent of an implication is not valid. Gentzen devises a method that uses numerical labels for a uniform treatment of iterated implications of any

complexity. The method is especially useful for disproving intuitionistic validity, as Gentzen shows.

In 1939, with the work for the *Habilitationsschrift* (Gentzen 1943) finished, Gentzen began two new series of notes, the *BG* and *BTJZ*. The first stands for "Buch über Grundlagenforschung." The notes were often written on pieces of paper like pages stripped out of an almanac or at the back of the "Luftwachezettel" that he filled during his military service close to Brunswick. (His task was to observe the time, type, direction, and number of aircraft that flew over). The idea with the book on foundations research was to write a popular exposition, "exciting like a detective story," as he put it on one note.

The signum *BTJZ* stands for *Beweistheorie der intuitionistischen Zahlentheorie*. It comes after the definitive work on the provable initial segments of transfinite induction, a work that used the classical sequent calculus *LK*. Now Gentzen turns to the intuitionistic version, with notes of which some hundred-odd pages have been preserved. The page numbers run from 133 (May 1939) to 275 (April 1943).

It would seem that a lot of progress can be made through the study of the Gentzen papers regarding his early work, especially structural proof theory in its two forms, natural deduction and sequent calculus, and the application of these calculi to the consistency problem of arithmetic. A remarkable addition to our understanding of the research into logic and foundations of mathematics in the 1930s could thus be the result. The later, long *BTJZ* series relates to these topics to which also Gentzen's published papers belong.

The biggest single series, judging from the page numbers that run beyond 339, was the series *WA*. Close to a hundred of these pages have been preserved, from between pages 77 (August 1938) and 254 (August 1943). We noted the existence of a thirteen-page summary of *WA*. This series is thus the last, and also the most enigmatic, of the series in the Gentzen papers. One gets from the transcribed pages the impression, through a suggestive terminology drawn from geography, of a strange country that Gentzen had created all to himself.

BIBLIOGRAPHY

[Ackermann, 1940] W. Ackermann. Zur Widerspruchsfreiheit der Zahlentheorie, *Mathematische Annalen*, vol. 117, pp. 162–194, 1940.

[Bernays, 1936] P. Bernays. *Logical Calculus*. Lectures at the Institute for Advanced Study, Princeton 1936.

[Bernays, 1936a] P. Bernays. Review of Gentzen (1936), *The Journal of Symbolic Logic*, vol. 1, p. 75, 1936.

[Bernays, 1944] P. Bernays. Review of Gentzen (1943), *The Journal of Symbolic Logic*, vol. 9, pp. 70–72, 1944.

[Bernays, 1945] P. Bernays. Review of Ketonen (1944), *The Journal of Symbolic Logic*, vol. 10, pp. 127–130, 1945.

[Bernays, 1965] P. Bernays. Betrachtungen zum Sequenzen-kalkul, in *Contributions to Logic and Methodology in Honor of J. M. Bochenski*, pp. 1–44, North-Holland, Amsterdam, 1965.

[Bernays, 1970] P. Bernays. On the original Gentzen consistency proof for number theory, in J. Myhill et al., eds, *Intuitionism and Proof Theory*, pp. 409–417, North-Holland, Amsterdam, 1970.

[Beth, 1947] E. Beth. Review of Popper (1947), *Mathematical Reviews*, MR0021924, 1947.

[Beth, 1955] E. Beth. Semantic entailment and formal derivability, *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen*, Afd. Letterkunde, vol. 18, pp. 309–342, 1955.

[Church, 1932] A. Church. A set of postulates for the foundation of logic, *Annals of Mathematics*, vol. 33, pp. 346–366, 1932.

[Church, 1941] A. Church. *The Calculi of Lambda-Conversion*, Princeton University Press.

[Curry, 1963] H. Curry. *Foundations of Mathematical Logic*, 1963. As republished by Dover, New York 1977.

[Curry and Feys, 1958] H. Curry and R. Feys. *Combinatory Logic*, vol. 1, North-Holland, Amsterdam, 1958.

[Dawson and Dawson, 2005] J. Dawson and C. Dawson. Future tasks for Gödel scholars, *The Bulletin of Symbolic Logic*, vol. 11, pp. 150–171, 2005.

[Dragalin, 1988] A. Dragalin. *Mathematical Intuitionism: Introduction to Proof Theory*, American Mathematical Society, Providence, Rhode Island, 1988. Russian original 1979.

[Dummett, 1975] M. Dummett. The philosophical basis of intuitionistic logic, 1975. As reprinted in Dummett, *Truth and Other Enigmas*, pp. 215–247, Duckworth, London 1978.

[Dyckhoff, 1992] R. Dyckhoff. Contraction-free sequent calculi for intuitionistic logic, *The Journal of Symbolic Logic*, vol. 57, pp. 795–807, 1992.

[Feferman, 1964] S. Feferman. Systems of predicative analysis, *The Journal of Symbolic Logic*, vol. 29, pp. 1–30, 1964.

[Fenstad, 1971] J. Fenstad, ed. *Proceedings of the Second Scandinavian Logic Symposium*, North-Holland, 1971.

[Gentzen, 1932] G. Gentzen. Über die Existenz unabhängiger Axiomensysteme zu unendlichen Satzsysteme, *Mathematische Annalen*, vol. 107, pp. 329–350, 1932.

[Gentzen, 1933] G. Gentzen. Über das Verhältnis zwischen intuitionistischer und klassischer Arithmetik, submitted for publication but withdrawn, published in *Archiv für mathematische Logik*, vol. 16 (1974), pp. 119-132, 1933.

[Gentzen, 1934–35] G. Gentzen. Untersuchungen über das logische Schliessen, *Mathematische Zeitschrift*, vol. 39, pp. 176-210 and 405-431, 1934–35.

[Gentzen, 1936] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie, *Mathematische Annalen*, vol. 112, pp. 493–565, 1936.

[Gentzen, 1936a] G. Gentzen. Die Widerspruchsfreiheit der Stufenlogik, *Mathematische Zeitschrift*, vol. 41, pp. 357–366, 1936.

[Gentzen, 1936–37] G. Gentzen. Der Unendlichkeitsbegriff in der Mathematik, *Semester-Berichte Münster*, Winter-Semester 1936–37, pp. 65–80. Reprinted in Menzler-Trott (2001), pp. 331–339, 1936–37.

[Gentzen, 1937] G. Gentzen. Unendlichkeitsbegriff und Widerspruchsfreiheit der Mathematik, *Actualités scientifiques et industrielles*, vol. 535, pp. 201–205, Hermann, Paris, 1937. Reprinted in Menzler-Trott (2001), pp. 340–342.

[Gentzen, 1938] G. Gentzen. Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie, *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften*, vol. 4, pp. 19–44, 1938.

[Gentzen, 1938a] G. Gentzen. Die gegenwärtige Lage in der mathematischen Grundlagenforschung, *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften*, vol. 4, pp. 5-18, 1938. Reprinted in Menzler-Trot (2001), pp. 343–358.

[Gentzen, 1943] G. Gentzen. Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie, *Mathematische Annalen*, vol. 120, pp. 140–161, 1943.

[Gentzen, 1954] G. Gentzen. Zusammenfassung von mehreren vollständigen Induktionen zu einer einzigen, *Archiv für mathematische Logik*, vol. 2, pp. 1–3, 1954.

[Gentzen, 1969] G. Gentzen. *The Collected Papers of Gerhard Gentzen*, ed. M. Szabo, North-Holland, Amsterdam, 1969.

[Gentzen, 1974] G. Gentzen. Der erste Widerspruchsfreiheitsbeweis für die klassische Zahlentheorie, *Archiv für mathematische Logik*, vol. 16, pp. 97–118, 1974.

[Girard, 1971] J.-Y. Girard. Une extension de l'interprétation de Gödel à l'analyse, et son application à l'elimination des coupures dans l'analyse et la théorie des types, in Fenstad (1971), pp. 63–92.

[Girard, 1987] J.-Y. Girard. *Proof Theory and Logical Complexity*, vol. 1, Bibliopolis, Naples, 1987.

[Glivenko, 1929] V. Glivenko. Sur quelques points de la logique de M. Brouwer, *Académie Royale de Belgique, Bulletins de la Classe des Sciences*, vol. 15, pp. 183–188, 1929.

[Gödel, 1931] K. Gödel. On formally undecidable propositions of *Principia mathematica* and related systems I, 1931. (Translation of German original), in van Heijenoort (1967), pp. 596–617. German original and translation also in Gödel (1986).

[Gödel, 1933] K. Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie, 1933. As reprinted in Gödel (1986), pp. 286–295.

[Gödel, 1938] K. Gödel. Lecture at Zilsel's, 1938. A lecture first published in Gödel (1995), pp. 86–113.

[Gödel, 1941] K. Gödel. In what sense is intuitionistic logic constructive?, 1941. A lecture first published in Gödel (1995), pp. 189–200.

[Gödel, 1958] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, vol. 12, pp. 280–287, 1958. Also in Gödel (1990), pp. 240–251.

[Gödel, 1986, 1990, 1995, 2003] K. Gödel. *Collected Works I–V*, Oxford University Press, 1986, 1990, 1995, 2003.

[van Heijenoort, 1967] J. van Heijenoort, ed. *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*, Harvard University Press, 1967.

[Hertz, 1929] P. Hertz. Über Axiomensysteme für beliebige Satzsysteme, *Mathematische Annalen*, vol. 101, pp. 457–514, 1929.

[Heyting, 1930] A. Heyting. Die formalen Regeln der intuitionistischen Logik, *Sitzungsberichte der Preussischen Akademie von Wissenschaften, Physikalisch-mathematische Klasse*, pp. 42–56, 1930.

[Heyting, 1931] A. Heyting. Die intuitionistische Grundlegung der Mathematik, *Erkenntnis*, vol. 2, pp. 106–115, 1931.

[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*, Springer, Berlin, 1928.

[Hilbert and Bernays, 1934, 1939] D. Hilbert and P. Bernays. *Grundlagen der Mathematik*, 2 vols., Springer, Berlin, 1934, 1939.

[Howard, 1969] W. Howard. The formulae-as-types notion of construction, 1969. Published in 1980 in J. Seldin and J. Hindley, eds, *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, pp. 480–490, Academic Press, New York.

[Hudelmaier, 1992] J. Hudelmaier. Bounds for cut elimination in intuitionistic propositional logic, *Archive for Mathematical Logic*, vol. 31, pp. 331–354, 1992.

[Jaskowski, 1934] S. Jaskowski. On the rules of supposition in formal logic, 1934. As reprinted in S. McCall, ed, *Polish Logic 1920–1939*, pp. 232–258, Oxford 1967.

[Johansson, 1936] I. Johansson. Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus, *Compositio Mathematica*, vol. 4, pp. 119–136, 1936.

[Kalmar, 1938] L. Kalmar. *Über Gentzens Beweis für die Widerspruchsfreiheit der reinen Zahlentheorie*, ms. dated 1938 IX.24, 21 pp., Paul Bernays archives, ETH-Zurich.

[Ketonen, 1943] O. Ketonen. "Luonnollisen päättelyn" kalkyylista, (On the calculus of "natural deduction," in Finnish), *Ajatus* (Yearbook of the Finnish Philosophical Society), vol. 12, pp. 128–140, 1943.

[Ketonen, 1944] O. Ketonen. *Untersuchungen zum Prädikatenkalkül*, (Annales Acad. Sci. Fenn, Ser. A.I. 23), 1944.

[Kleene, 1952] S. Kleene. *Introduction to Metamathematics*, North-Holland, Amsterdam, 1952.

[Kleene, 1952a] S. Kleene. Permutability of inferences in Gentzen's calculi LK and LJ, *Memoirs of the American Mathematical Society*, No. 10, pp. 1–26, 1952.

[Kolmogorov, 1925] A. Kolmogorov. On the principle of excluded middle, 1925.(Translation of Russian original), in van Heijenoort (1967), pp. 416–437.

[Kolmogorov, 1932] A. Kolmogorov. Zur Deutung der intuitionistischen Logik, *Mathematische Zeitschrift*, vol. 35, pp. 58–65, 1932.

[Kreisel, 1987] G. Kreisel. Gödel's excursions into intuitionistic logic, in P. Weingartner and L. Schmetterer, eds, *Gödel Remembered*, pp. 65–186, Bibliopolis, Naples, 1987.

[Martin-Löf, 1975] P. Martin-Löf. An intuitionistic theory of types: predicative part, in H. Rose and J. Shepherson, eds, *Logic Colloquium '73*, pp. 73–118, North-Holland, Amsterdam, 1975.

[Martin-Löf, 1984] P. Martin-Löf. *Intuitionistic Type Theory*, Bibliopolis, Naples, 1984.

[Menzler-Trott, 2001] E. Menzler-Trott. *Gentzens Problem: mathematische Logik im nationalsozialistischen Deutschland*, Birkhäuser, Basel, 2001.

[Menzler-Trott, 2007] E. Menzler-Trott. *Logic's Lost Genius* American Mathematical Society, Providence, Rhode Island. 2007.

[Negri, 1999] S. Negri. Sequent calculus proof theory of intuitionistic apartness and order relations, *Archive for Mathematical Logic*, vol. 38, pp. 521–547, 1999.

[Negri, 2003] S. Negri. Contraction-free sequent calculi for geometric theories with an application to Barr's theorem, *Archive for Mathematical Logic*, vol. 42, pp. 389–401, 2003.

[Negri and von Plato, 1998] S. Negri and J. von Plato. Cut elimination in the presence of axioms, *The Bulletin of Symbolic Logic*, vol. 4 , pp. 418–435, 1998.

[Negri and von Plato, 2001] S. Negri and J. von Plato. *Structural Proof Theory*, Cambridge University Press, 2001.

[Negri and von Plato, 2001a] S. Negri J. von Plato. Sequent calculus in natural deduction style, *The Journal of Symbolic Logic*, vol. 66, pp. 1803–1816, 2001.

[Negri and von Plato, 2004] S. Negri and J. von Plato. Proof systems for lattice theory, *Mathematical Structures in Computer Science*, vol. 14, pp. 507–526, 2004.

[von Plato, 2001] J. von Plato. A proof of Gentzen's *Hauptsatz* without multicut, *Archive for Mathematical Logic*, vol. 40, pp. 9–18, 2001.

[von Plato, 2001a] J. von Plato. Natural deduction with general elimination rules, *Archive for Mathematical Logic*, vol. 40, pp. 541–567, 2001.

[von Plato, 2003] J. von Plato. Translations from natural deduction to sequent calculus, *Mathematical Logic Quarterly*, vol. 49, pp. 435-443, 2003.

[von Plato, 2004] J. von Plato. The decision problem in projective and affine geometry, to appear, 2004.

[von Plato, 2004a] J. von Plato. Ein Leben, ein Werk – Gedanken über das wissenschaftliche Schaffen des finnischen Logikers Oiva Ketonen, in R. Seising, ed, *Form, Zahl, Ordnung: Studien zur Wissenschafts- und Technikgeschichte*, pp. 427–435, Franz Steiner Verlag, Stuttgart, 2004.

[von Plato, 2005] J. von Plato. Gentzen's original proof of the consistency of arithmetic revisited, to appear, 2005.

[Popper, 1947] K. Popper. New foundations for logic, *Mind*, vol. 56, pp. 193–235, 1947.

[Post, 1936] E. Post. Finite combinatory processes – formulation 1, *The Journal of Symbolic Logic*, vol. 1, pp. 103–105, 1936.

[Prawitz, 1965] D. Prawitz. *Natural Deduction: A Proof-Theoretical Study*. Almqvist & Wicksell, Stockholm, 1965.

[Prawitz, 1971] D. Prawitz. Ideas and results in proof theory, in Fenstad (1971), pp. 235–308, 1971.

[Raggio, 1965] A. Raggio. Gentzen's Hauptsatz for the systems NI and NK, *Logique et analyse*, vol. 8, pp. 91–100, 1965.

[Rathjen, 1995] M. Rathjen. Recent advances in ordinal analysis, *The Bulletin of Symbolic Logic*, vol. 1, pp. 468–485, 1995.

[Rathjen, 1999] M. Rathjen. The realm of ordinal analysis, in S. Wainer and J. Truss, eds, *Sets and Proofs*, pp. 219–279, Cambridge University Press, 1999.

[Reid, 1970] C. Reid. *Hilbert*, Springer, 1970.

[Rosser, 1937] B. Rosser. Gödels theorems for non-constructive logics, *The Journal of Symbolic Logic*, vol. 2, pp. 129–137, 1937.

[Schmidt, 1935] A. Schmidt. Review of Gentzen (1934–35), *Zentralblatt für Mathematik und ihre Grenzgebiete*, vol. 10, pp. 145–146, 1935.

[Schmidt, 1936] A. Schmidt. Review of Gentzen (1936a), *Zentralblatt für Mathematik und ihre Grenzgebiete*, vol. 15, p. 193, 1936.

[Schroeder-Heister, 1984] P. Schroeder-Heister. A natural extension of natural deduction, *The Journal of Symbolic Logic*, vol. 49, pp. 1284–1300, 1984.

[Schroeder-Heister, 2002] P. Schroeder-Heister. Resolution and the origins of structural reasoning: early proof-theoretic ideas of Hertz and Gentzen, *The Bulletin of Symbolic Logic*, vol. 8, pp. 246–265, 2002.

[Schütte, 1950] K. Schütte. Schlussweisen-Kalküle der Prädikatenlogik, *Mathematische Annalen*, vol. 122, pp. 47–65, 1950.

[Schütte, 1965] K. Schütte. Eine Grenze für die Beweisbarkeit der transfiniten Induktion in der verzweigten Typenlogik, *Archiv für mathematische Logik*, vol. 7, pp. 45–60, 1965.

[Skolem, 1920] T. Skolem. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen, 1920. As reprinted in Skolem's *Selected Works in Logic*, pp. 103–136, ed. Fenstad, Universitetsforlaget, Oslo 1970.

[Tait, 1967] W. Tait. Intensional interpretations of functionals of finite type I, *The Journal of Symbolic Logic*, vol. 32, pp. 198–212, 1967.

[Tait, 2001] W. Tait. Gödel's unpublished papers on foundations of mathematics, *Philosophia Mathematica*, vol. 9, pp. 87–126, 2001.

[Tait, 2005] W. Tait. Gödel's reformulation of Gentzen's first consistency proof for arithmetic: the no-counterexample interpretation, *The Bulletin of Symbolic Logic*, vol. 11, pp. 225–238, 2005.

[Takeuti, 1975] G. Takeuti. *Proof Theory*, North-Holland, Amsterdam, 1975.

[Tarski, 1933] A. Tarski. Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit, *Monatshefte für Mathematik und Physik*, vol. 40, pp. 97–112, 1933.

[Troelstra, 1986] A. Troelstra. Comments to Gödel (1933), in Gödel (1986), p. 282.

[Troelstra and Schwichtenberg, 2000] A. Troelstra and H. Schwichtenberg. *Basic Proof Theory*, second ed., Cambridge, 2000. First ed. 1996.

LAMBDA-CALCULUS AND COMBINATORS IN THE 20TH CENTURY

Felice Cardone and J. Roger Hindley

1 INTRODUCTION

The formal systems that are nowadays called λ-calculus and combinatory logic were both invented in the 1920s, and their aim was to describe the most basic properties of function-abstraction, application and substitution in a very general setting. In λ-calculus the concept of abstraction was taken as primitive, but in combinatory logic it was defined in terms of certain primitive operators called *basic combinators*.

The present article will sketch the history of these two topics, until approximately the year 2000.

We shall assume the reader is familiar with at least one of the many versions of these systems in the current literature. A few key technical details will be given as footnotes, however.[1] Often "combinatory logic" will be abbreviated to "CL" and "λ-calculus" to "λ". We shall distinguish between "pure" versions of λ or CL (theories of conversion or reduction with nothing more) and "applied" versions (containing extra concepts such as logical constants, types or numbers).

To understand the early history it is worth remembering that the situation in logic and the foundations of mathematics was much more fluid in the early 1900s than it is today; Russell's paradox was relatively recent, Gödel's theorems were not yet known, and a significant strand of work in logic was the building of systems intended to be consistent foundations for the whole of mathematical analysis. Some of these were based on a concept of set, others on one of function, and there was no general consensus as to which basis was better. In this context λ and CL were originally developed, not as autonomous systems but as parts of more elaborate foundational systems based on a concept of function.

Today, λ and CL are used extensively in higher-order logic and computing. Rather like the chassis of a bus, which supports the vehicle but is unseen by

[1] A short introduction to λ-calculus is included in [Seldin, 2008] in the present volume. Others can be found in many textbooks on computer science. There are longer introductions in [Hankin, 1994], [Hindley and Seldin, 1986], [Stenlund, 1972]; also in [Krivine, 1990] (in French and English), [Takahashi, 1991] (in Japanese), and [Wolfengagen, 2004] (in Russian). A deeper account is in [Barendregt, 1981]. For combinatory logic there are introductions in [Hindley and Seldin, 1986, Ch.2], [Stenlund, 1972], and [Barendregt, 1981, Ch.7]; more detailed accounts are in [Curry and Feys, 1958, Chs.5–9] and [Curry *et al.*, 1972].

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

its users, versions of λ or CL underpin several important logical systems and programming languages. Further, λ and CL gain most of their purpose at second hand from such systems, just as an isolated chassis has little purpose in itself. Therefore, to give a balanced picture the present article should really include the whole history of function-based higher-order logic. However, it would then be too diffuse, so we shall take a more restricted approach. The reader should always keep the wider context in mind, however.

Seen in outline, the history of λ and CL splits into three main periods: first, several years of intensive and very fruitful study in the 1920s and '30s; next, a middle period of nearly 30 years of relative quiet; then in the late 1960s an upsurge of activity stimulated by developments in higher-order function theory, by connections with programming languages, and by new technical discoveries. The fruits of the first period included the first-ever proof that predicate logic is undecidable. The results of the second attracted very little non-specialist interest, but included completeness, cut-elimination and standardization theorems (for example) that found many uses later. The achievements of the third, from the 1960s onward, included constructions and analyses of models, development of polymorphic type systems, deep analyses of the reduction process, and many others probably well known to the reader. The high level of activity of this period continues today.

The present article will describe earlier work in chronological order, but will classify later developments by topic, insofar as overlaps allow. Each later topic will be discussed in some breadth, to show something of the contexts in which λ and CL are being used. However, due to lack of space and the richness of the field we shall not be able to be as comprehensive as we would like, and some important sub-topics will unfortunately have to be omitted; for this we ask the reader's forgiveness. Although we shall try to keep a balance, our selection will inevitably reflect our own experience. For example Curry's work will be given more space than Church's simply because we know it better, not because we think it more important.

A large bibliography will be included for the reader's convenience in tracing sources. When detailed evidence of a source is needed, the relevant precise section or pages will be given. Sometimes the date we shall give for an idea will significantly precede its date of publication; this will be based on evidence in the source paper such as the date of the manuscript. By the way, mention of a date and an author for an idea should not be interpreted as a claim of priority for that author; important ideas may be invented independently almost simultaneously, and key papers may be circulated informally for some years before being published, so questions of priority, although very interesting, are beyond our powers to decide.

The bibliography will be nowhere near complete, nor intended to be; space will not allow that, and two very comprehensive bibliographies on λ and CL have already been published: [Rezus, 1982], covering up to 1980, and [Bethke, 2000].

Acknowledgements

The present account owes its beginning to a suggestion by Dirk van Dalen.

For useful information and advice during its preparation, the authors are very grateful to many helpful correspondents, in particular John Addison, Peter Andrews, Corrado Böhm, Martin Bunder, Thierry Coquand, Pierre-Louis Curien, Steven Givant, Ivor Grattan-Guinness, Peter Hancock, Gérard Huet, Reinhard Kahle, Jan-Willem Klop, Alexander Kuzichev, William Lawvere, Bruce Lercher, Giuseppe Longo, Per Martin-Löf, Eugenio Moggi, John Reynolds, Simona Ronchi della Rocca, Jonathan Seldin, John Shepherdson, William Tait, Christian Thiel, Anne Troelstra, Pawel Urzyczyn and Philip Wadler.

This account has also been much helped by material from other historical articles and books, especially [Barendregt, 1997], [Crossley, 1975], [Curry and Feys, 1958, §§0D, 1S, 2S, 3S, etc.], [Gandy, 1988], [Grattan-Guinness, 2000], [Heijenoort, 1967], [Hodges, 1983], [Kalman, 1983], [Kamareddine *et al.*, 2002], [Kleene, 1981], [Laan, 1997], [Manzano, 1997], [Nederpelt and Geuvers, 1994], [Rosser, 1984], [Seldin, 1980a; Seldin, 1980b], and [Sieg, 1997].

However, any errors in what follows are our own responsibility.

On the financial side, we express our gratitude to the Dipartimento di Informatica of the University of Turin, and the British Council, for their very generous support and facilities which made a crucial consultation visit possible in May 2000.

2 PRE-HISTORY

Notations for function-abstraction and substitution go back at least as far as 1889. In Giuseppe Peano's book on axioms for arithmetic [Peano, 1889, §VI], for any term α containing a variable x, the function of x determined by α was called $\alpha[x]$, and for $\phi = \alpha[x]$ the equation $\phi\, x' = \alpha[x]x'$ was given, and its right-hand side was explicitly stated to mean the result of substituting x' for x in α. Later, Peano used other function-abstraction notations instead of $\alpha[x]$; these included $\alpha\,\overline{x}$ and $\alpha|x$, in [Peano, 1958, p.277] and [Peano, 1895, Vol.3 §11] respectively.

In 1891 Gottlob Frege discussed the general concept of function and introduced the notion of a function as a graph (*Wertverlauf* in [Frege, 1891]). Two years later, notations for abstraction and application appeared in his [Frege, 1893, Vol.1 §9], where the graph of a function $\Phi(\ )$ was denoted by $\dot{\varepsilon}\,\Phi(\varepsilon)$, and the result of applying the function to an argument Δ was denoted by $\Delta \cap \dot{\varepsilon}\,\Phi(\varepsilon)$. In the same work Frege also proved the equation $a \cap \dot{\varepsilon}\, f(\varepsilon) = f(a)$, [Frege, 1893, Vol.1 p.73, §54].

Notations for class-abstraction too appeared at around the same time. The set of all x satisfying condition α was denoted by $[x\,\varepsilon]\,\alpha$ in [Peano, 1889, §V], and by $\overline{x\,\varepsilon}\,\alpha$ in Peano's work from 1890 onward, see [Peano, 1958, Vol.2 p.110]. The latter notation was also used by his colleague Burali-Forti [Burali-Forti, 1894, p.71 §3]. Another class-abstraction notation, $\hat{z}(\phi\, z)$, appeared in [Russell and Whitehead, 1913, Ch.1, p.25]. The latter two authors also used $\varphi\,\hat{x}$ for the unary propositional function determined by $\varphi\, x$.

A function-abstraction notation was used by Bertrand Russell in 1903–5 in unpublished notes, see [Klement, 2003].

However, none of these authors gave formal definitions of substitution and conversion, which are characteristic of CL and λ proper. These did not appear until the 1920s, and, although λ might seem to express the concepts involved more directly than CL, it was actually CL that was invented first.

3 1920s: BIRTH OF COMBINATORY LOGIC

Combinatory logic was invented by Moses Ilyich Schönfinkel. Schönfinkel was born in 1887 or '89 in Dniepropetrovsk in the Ukraine, and studied under the Odessa mathematician Samuel Shatunovsky – a brilliant pupil, according to [Yanovskaya, 1948, p. 31]. From 1914 to 1924 he belonged to what was probably the top mathematics research group of the time, that led by David Hilbert in Göttingen, Germany. He introduced combinators to this group in a talk on 7 December 1920, which was published in [Schönfinkel, 1924].

Schönfinkel's main aim was to eliminate bound variables, which he saw as merely auxiliary syntactic concepts, see [Schönfinkel, 1924, end of §1]. To do this, he introduced five operators Z, T, I, C and S which would nowadays be called *basic combinators* $\mathbf{B}$, $\mathbf{C}$, $\mathbf{I}$, $\mathbf{K}$ and $\mathbf{S}$ respectively.[2]

Schönfinkel proved that $\mathbf{K}$ and $\mathbf{S}$ are adequate to define the other three basic combinators, and stated a form of combinatory completeness result (with a sketch proof): namely that from $\mathbf{K}$ and $\mathbf{S}$, together with a logical operator equivalent to a universally quantified Sheffer stroke or *Nand*, one can generate all formulas of predicate logic without the use of bound variables. But he treated equality and logical equivalence only informally, as was common practice in mathematics at that time; he did not introduce formal conversion rules. Nor was an explicit abstraction algorithm stated in his paper, although from what was written it seems extremely likely that he knew one.

Along the way, Schönfinkel pointed out that multi-variable applications such as $F(x, y)$ could be replaced by successive single applications $(f(x))(y)$, where f was a function whose output-value $f(x)$ was also a function, see [Schönfinkel, 1924, §2]. This replacement-process is now known in the computer-science community as "currying", from its use in the work of Haskell Curry; although Curry many times attributed it to Schönfinkel, for example in [Curry, 1930, p.512] and [Curry and Feys, 1958, pp. 8, 11, 30, 106], and it was used even earlier by Frege, [Frege, 1893, Vol.1 §4].

[2] In today's notation (following [Curry and Feys, 1958, Ch.5]) their axiom-schemes are $\mathbf{B}XYZ = X(YZ)$, $\mathbf{C}XYZ = XZY$, $\mathbf{I}X = X$, $\mathbf{K}XY = X$, $\mathbf{S}XYZ = XZ(YZ)$, and a *combinator* is any applicative combination of basic combinators. A set $\mathcal{S}$ of combinators is called *combinatorially complete* when, for every sequence $x_1, ..., x_n$ of variables and every combination X of all or some of these (possibly with some occurring more than once), there exists a combination A of members of $\mathcal{S}$ such that $Ax_1...x_n = X$. An *abstraction algorithm* is any algorithm for constructing a suitable A from X. In what follows, the notation "$[x_1, ..., x_n].X$" will denote any suitable A.

After his 1924 paper, Schönfinkel published nothing more on combinators. In fact only one other paper bears his name: [Bernays and Schönfinkel, 1928], on certain special cases of the *Entscheidungsproblem*. Both papers were prepared for publication largely by helpful colleagues: the earlier one by Heinrich Behmann and the later by Paul Bernays. By 1927, he was said to be mentally ill and in a sanatorium [Curry, 1927, p.3], [Kline, 1951, p.47]. Alexander Kuzichev in Moscow relates that Schönfinkel's later years were spent in Moscow in hardship and poverty, helped by a few friends, and he died in a hospital there in 1942 following some years of illness. After his death, wartime conditions forced his neighbours to burn his manuscripts for heating.[3]

However, in the meantime the combinator idea had appeared again, though in a very modified form. In 1925, John von Neumann published his doctoral thesis on axiomatizing set theory, [Neumann, 1925]. Despite its being a theory of sets, his formal system was function-based not set-based, and one group of his axioms postulated certain combinator-like operators which gave combinatory completeness for all expressions built from variables and constants by pairing and function-application, [Neumann, 1925, p.225].[4]

(We do not know whether von Neumann's idea came from Schönfinkel's. Von Neumann visited Göttingen well before 1925 and collaborated with the logic group there, and it is hard to imagine that he did not learn of Schönfinkel's work there. But the form in which the concept appeared in [Neumann, 1925] was very modified, and his [Neumann, 1925] and [Neumann, 1928] did not mention Schönfinkel.)

The von Neumann axioms eventually evolved, with many changes, including from a function base to a sets-and-classes base, into the nowadays well known Von-Neumann-Bernays-Gödel system NBG. In that system the analogue of von Neumann's combinator axioms was a finite list of class-existence axioms, and the analogue of his combinatory completeness theorem was a general class-existence theorem. That system gave a finite axiomatization of set theory, in contrast to the Zermelo-Fraenkel system known nowadays as ZF. It did not, however, eliminate bound variables completely from the logic; unlike Schönfinkel, von Neumann did not have that as an aim.

The next major step in CL was independent of both Schönfinkel and von Neu-

[3]Little more is known about him. We are indebted mainly to [Thiel, 1995], [Yanovskaya, 1948, pp.31–34], [Kline, 1951], and correspondence from A. S. Kuzichev, 2005. His date of birth was 4th Sept. 1889 according to [Thiel, 1995], but 1887 according to Kuzichev.

[4]The relevant axioms appeared in [Neumann, 1925, §3, axiom-group II], see comments in [Curry and Feys, 1958, pp.10–11]. We show them here in a notation chosen to emphasize their similarity to CL. Von Neumann used two atomic predicates, "x is a function" and "x is an argument" (not mutually exclusive), and two term-building operations: pairing, which we call "$\langle\,,\,\rangle$", and application "$(\ )$". (For pairing, he postulated two projection-functions; we omit them here.)

$(\exists\, \mathit{fn}\ \mathsf{I})(\forall\ \mathit{args}\ x)$	$(\mathsf{I}x) = x;$
$(\forall\ \mathit{args}\ x)(\exists\, \mathit{fn}\ \mathsf{K}_x)(\forall\ \mathit{args}\ y)$	$(\mathsf{K}_x y) = x;$
$(\exists\, \mathit{fn}\ \mathbf{App})(\forall\ \mathit{args}\ f, x)$	$f\ \mathit{is\ a\ fn}\ \supset\ (\mathbf{App}\,\langle f, x\rangle) = (fx);$
$(\forall\ \mathit{fns}\ f, g)(\exists\, \mathit{fns}\ \mathsf{S}_{f,g}, \mathsf{B}_{f,g})(\forall\ \mathit{args}\ x)$	$(\mathsf{S}_{f,g}\, x) = \langle (fx), (gx)\rangle\ \wedge\ (\mathsf{B}_{f,g}\, x) = (f(gx)).$

mann: the combinator concept was re-invented in the U.S.A. by Haskell Curry.

Curry was born in 1900. His parents ran a school of public speaking near Boston, U.S.A. (now called Curry College), and from them he derived an interest in linguistics which coloured much of his later work. He began his student career at Harvard University intending to work in medicine but quickly changed to mathematics, and a strong underlying inclination towards foundational questions drew him eventually to logic.[5]

In propositional logic the rule of substitution is significantly more complex than *modus ponens* and the other deduction-rules, and at the time Curry began his studies the substitution operation was largely unanalyzed.[6] In particular, an explicit statement of the substitution rule was missing from *Principia Mathematica* [Russell and Whitehead, 1913], as was admitted later by Russell himself, see [Russell, 1919, Ch.14, p.151]. Also the corresponding rule for predicate logic was stated incorrectly in the first edition of a leading textbook [Hilbert and Ackermann, 1928, Ch.3 §5 p.53], see remarks in [Church, 1956, §49 pp. 290–291].

Around 1926–27, Curry began to look for a way to break the substitution process down into simpler steps, and to do this, he introduced a concept of combinator essentially the same as Schönfinkel's. Then in late 1927 he discovered [Schönfinkel, 1924], and a note in his files [Curry, 1927, p.1] admits "This paper anticipates much of what I have done". He did not abandon his own work, however, but travelled to Göttingen and in 1930 completed a doctoral thesis under the guidance of Bernays, [Curry, 1930]. In it he set out a formal system of combinators with a proof of the combinatory completeness of $\{\mathbf{B}, \mathbf{C}, \mathbf{K}, \mathbf{W}\}$ [Curry, 1930, p.528 Hauptsatz II].[7]

Crucial to every such completeness proof is an abstraction algorithm. The one in Curry's thesis was not only the first to appear in print, but was also much more efficient than many that followed it. He saw the problem of producing an A such that $Ax_1...x_n$ converted to X, from the point of view of building X from the sequence $x_1, \ldots, x_n$. First, A must remove from $x_1, \ldots, x_n$ all variables which do not occur in X; this can be done using $\mathbf{K}$ and $\mathbf{B}$. Then A must repeat variables as often as they occur in X; this can be done using $\mathbf{W}$ and $\mathbf{B}$. Next, A must re-arrange the variables into the order in which they occur in X; this can be done using $\mathbf{C}$ and $\mathbf{B}$. Finally, A must insert the parentheses that occur in X, and this can be done using $\mathbf{B}$ (together with $\mathbf{I}$ or $\mathbf{CKK}$).[8]

But the details of Curry's algorithm were rather complex, and it was later replaced by ones that were much simpler to describe. These algorithms were multi-sweep; i.e. they built $[x_1, \ldots, x_n].X$ by building first $[x_n].X$, then $[x_{n-1}].[x_n].X$, etc., and their key was to build $[x].X$ for all x and X by a direct and very simple

[5] For information on Curry we are indebted to [Curry and Feys, 1958, §§1S,6S,8S], [Curry, 1980], and [Seldin, 1980b; Seldin, 1980a], also to discussions with Jonathan Seldin, whose [Seldin, 2008] gives a fuller account of Curry and Church than the present one. Curry died in 1982.

[6] The substitution rule in propositional logic states that from a formula Q we may deduce $[P_1/p_1, \ldots, P_n/p_n]Q$, where the latter is obtained from Q by substituting formulas $P_1, \ldots, P_n$ for propositional variables $p_1, \ldots, p_n$ which may occur in Q.

[7] $\mathbf{W}$ has the axiom-scheme $\mathbf{W}XY = XYY$.

[8] See [Curry, 1930, Part II, §§C–E] for details, or [Curry and Feys, 1958, §6S2] for an outline.

induction on the number of symbols in X.

The first to appear was in [Church, 1935, §3 p.278]. Church used just two basic combinators I and J, where $\mathsf{J}UXYZ = UX(UZY)$; the latter had been proposed by Rosser in 1933 [Rosser, 1935, p.128].[9]

But the simple algorithm based on S and K that is used in textbooks today is due to Paul Rosenbloom [Rosenbloom, 1950, p.117]. He seems to have been the first person, at least in print, to have realised that by using S in the induction step for $[x].X$ the proof of its main property $([x].X)Y = [Y/x]X$ was made trivial.[10] This algorithm, with some variants, became generally accepted as standard.[11]

However, although neater to describe, the direct inductive algorithms were not so efficient to use: they produced considerably longer outputs than Curry's first one, and in the 1970s when the desire arose to use combinators in practical programming languages, and attention began to focus on efficiency, the algorithms that were invented then were more like Curry's original, see for example [Kearns, 1973; Abdali, 1976; Turner, 1979; Piperno, 1989].

Besides the first abstraction algorithm, Curry's thesis contained the first formal definition of conversion. Further, this definition included a finite set of axioms from which he proved the admissibility of rule (ζ):[12]

(ζ) *if* $Ux = Vx$ *and* x *does not occur in* UV, *then* $U = V$.

Thus the discovery that (ζ) could be finitely axiomatized was made very soon after the birth of the theory. The theory of equality was also proved consistent in a certain sense, see [Curry, 1930, p.528 Hauptsatz I], but without using a concept of strong reduction or a confluence theorem.

The system in [Curry, 1930] also contained some logical constants. But not all their usual rules were included, negation was omitted, and logic was not actually

[9]The basis $\{\mathsf{I}, \mathsf{J}\}$ allows $[x].X$ to be built when x occurs in X but not when x is absent. This corresponds to the λI-calculus, see p.731 below. Church's algorithm is also in [Church, 1941, p.44].

[10]No he wasn't. [Seldin, 2008] points out that Curry realized this neat use of S while reading [Rosser, 1942], though, due to war work, he did not publish it until [Curry, 1949, p.393]. Thus the algorithm's direct inductive form is due to Church and its neat use of S to Curry and Rosenbloom independently.

[11]Some $[x].X$-algorithms were compared in [Curry and Feys, 1958, §§6A2–3]. Curry assumed the basic combinators were some of B, C, I, K, S, and $[x].X$ was built by induction on X using some or all of the following clauses (in which U does not contain x, but V may):

(a)	$[x].U$	$\equiv \mathsf{K}U$,	(d)	$[x].UV$	$\equiv \mathsf{B}U([x].V)$,
(b)	$[x].x$	$\equiv \mathsf{I}$,	(e)	$[x].VU$	$\equiv \mathsf{C}([x].V)U$,
(c)	$[x].Ux$	$\equiv U$,	(f)	$[x].X_1X_2$	$\equiv \mathsf{S}([x].X_1)([x].X_2)$.

For example, let the basis be $\{\mathsf{I}, \mathsf{K}, \mathsf{S}\}$. Then algorithm (fab) builds $[x].X$ by first applying (f) whenever possible, then (a), then (b) when the other two do not apply. Alternatively, (abcf) first tries to apply (a), then (b), then (c), and (f) only when the other three do not apply; it usually produces shorter terms $[x].X$. The Curry-Rosenbloom algorithm was (fab).

If the basis is $\{\mathsf{K}, \mathsf{S}\}$ one can first define $\mathsf{I} \equiv \mathsf{SKK}$ and then use (fab) or (abcf). If the basis is $\{\mathsf{B}, \mathsf{C}, \mathsf{K}, \mathsf{W}\}$ one can define $\mathsf{S} \equiv \mathsf{B}(\mathsf{B}(\mathsf{BW})\mathsf{C})(\mathsf{BB})$ and $\mathsf{I} \equiv \mathsf{CKK}$ or WK, and then use (abcdef).

[12]See [Curry, 1930, p.832 Satz 4] or [Curry, 1932, p.558 Theorem 5]. Rule (ζ) makes Curry's equality equivalent to what is nowadays called $\beta\eta$-equality or *extensional equality*.

developed beyond the theory of conversion. In fact, unlike von Neumann, Curry aimed, not at quickly building a system strong enough to deduce a significant part of mathematics, but at making a thorough study of basic logical notions in as general a setting as possible.

A consequence of this interest in generality was that for Curry, as for Schönfinkel, every combinator was allowed to be applied to every other combinator and even to itself, in a way not commonly accepted for set-theoretic functions. This freedom was later criticized on semantic grounds by several other leading logicians, e.g. Willard V. Quine in [Quine, 1936a, p.88]. In fact Curry's use of combinators was connected very closely to his philosophy of mathematics: for him, a formal system was not a description of some pre-existing objects, but simply a structure built by induction from some primitives using certain rules, and he did not demand that a semantics for such a system be a total function. In his view a system could be used fruitfully in applications of mathematics without necessarily every part of it having a meaning in some set-theory-based semantics, and usefulness was more important than semantic totality.[13]

In [Curry, 1964] he pointed out that the combinators constructed by his original abstraction algorithm were much easier to interpret in a set-theoretic semantics than those built by simpler algorithms using **K** and **S**. (But, as mentioned above, he himself was perfectly happy to use the simpler algorithms.)

However, semantic questions apart, by the end of the 1920s the combinator concept had provided two useful formal techniques: a computationally efficient way of avoiding bound variables, and a finite axiomatization of set theory.

4 1930s: BIRTH OF λ AND YOUTH OF CL

4.1 Early λ-calculus

The λ-calculus was invented in about 1928 by Alonzo Church, and was first published in [Church, 1932]. Church was born in 1903 in Washington D.C. and studied at Princeton University. He made his career at Princeton until 1967, though in 1928–29 he visited Göttingen and Amsterdam.[14]

Around 1928 he began to build a formal system with the aim of providing a foundation for logic which would be more natural than Russell's type theory or Zermelo's set theory, and would not contain free variables (for reasons he explained in [Church, 1932, pp. 346–347]). He chose to base it on the concept of function rather than set, and his primitives included abstraction $\lambda x[M]$ and application $\{F\}(X)$, which we shall call here "$\lambda x.M$" and "(FX)".

[13] See [Curry, 1951], [Curry, 1963, Chs. 2–3] or [Seldin, 1980a]. In the 1920s the general concept of formal system was still in the process of clarification and Curry was one of the pioneers in this.

[14] For more on Church's life, see [Manzano, 1997, §2] and the obituary [Enderton, 1995]. For more on his work, see [Manzano, 1997], [Anderson, 1998], [Enderton, 1998], [Sieg, 1997], and (on the impact of λ-calculus) [Barendregt, 1997]. He died in 1995.

(By the way, why did Church choose the notation "λ"? In [Church, 1964, §2] he stated clearly that it came from the notation "$\hat{x}$" used for class-abstraction by Whitehead and Russell, by first modifying "$\hat{x}$" to "$\wedge x$" to distinguish function-abstraction from class-abstraction, and then changing "$\wedge$" to "λ" for ease of printing. This origin was also reported in [Rosser, 1984, p.338]. On the other hand, in his later years Church told two enquirers that the choice was more accidental: a symbol was needed and "λ" just happened to be chosen.)

As mentioned earlier, Church was not the first to introduce an explicit notation for function-abstraction. But he was the first to state explicit formal conversion rules for the notation, and to analyse their consequences in depth. (He did not read Frege's work until 1935, according to a letter cited in [Scott, 1980a, p.260], and we assume he did not see Peano's 1889 notation either.)

Church's system as it appeared in [Church, 1932] was a type-free logic with unrestricted quantification but without the law of excluded middle. Explicit formal rules of λ-conversion were included. However, almost immediately after publication a contradiction was found in it. The system was revised a year later, [Church, 1933], and in the revised paper Church stated a hope that Gödel's recent incompleteness theorems for the system of [Russell and Whitehead, 1913] did not extend to his own revised system, and that a finitary proof of the consistency of this system might be found, [Church, 1933, pp. 842–843]. At that time the full power of Gödel's theorems was only slowly becoming generally understood.

At the end of his 1933 paper, Church introduced the representation of the positive integers by the λ-terms now known as the *Church numerals* [Church, 1933, §9]:

$$\mathbf{1} \equiv \lambda xy.\, xy, \quad \mathbf{Succ} \equiv \lambda xyz.\, y(xyz), \quad \mathbf{n} =_\beta \lambda xy.\, \underbrace{x(\cdots(x}_{n \text{ times}}\, y)\cdots).$$

He defined the set of all positive integers in essentially the same way as the set of all finite cardinal numbers in [Russell, 1903, §123], except that Russell's set included zero. The natural λ-term to represent zero would have been $\lambda xy.y$, but although Church included this term in his language, he forbade the reduction $(\lambda xy.y)FY \triangleright Y$ and avoided zero.

(Church's name is often associated with the λI-*calculus*, the version of λ-calculus in which $\lambda x.M$ is only counted as a term when x occurs free in M. But he did not limit himself so strictly to this version as is often thought. In 1932 and '33 he allowed non-λI-terms $\lambda x.M$ to exist but not to be "active", i.e. he did not allow a term $(\lambda x.M)N$ to be contracted when x did not occur free in M, see [Church, 1932, pp. 352, 355]. In his 1940 paper on type theory, where consistency would have been less in doubt, although probably not yet actually proved, he was happy for non-λI-terms to be active [Church, 1940, pp. 57, 60]. Only in his 1941 book did he forbid their very existence [Church, 1941, p. 8].)[15]

From 1931 to 1934 in Princeton University, Church received the help of two outstanding graduate students, Stephen Kleene and Barkley Rosser, and in a re-

[15]The unrestricted λ-calculus is often called the λK-*calculus*.

markable four years of collaboration this group made a series of major discoveries about both Church's 1933 logic and the underlying pure λ-calculus. Unfortunately for the 1933 logic, one of these discoveries was its inconsistency, as a variant of the Richard paradox was proved in the system, [Kleene and Rosser, 1935].

In contrast, the pure λ-calculus, which had at first seemed shallow, turned out to be surprisingly rich. (Well, Kleene was surprised, anyway, [Kleene, 1981, p.54].) The study of its reduction-theory was begun using the method of residuals, which was applied to prove the Church-Rosser confluence theorem [Church and Rosser, 1936, p.479], and this ensured the consistency of the pure system. The system's relation to combinatory logic was also clarified, and a simple proof of the confluence of combinatory weak reduction was given; see [Rosser, 1935, esp. pp. 145–146].[16]

Further, the λ-definable numerical functions were found to form a much more extensive class than expected, and this class was eventually proved equivalent to two definitions whose purpose was to formalize the informal concept of effectively calculable function: the Herbrand-Gödel recursive functions, see [Kleene, 1936] and [Church, 1936b], and the Turing computable functions, [Turing, 1937a]. In view of this, Church conjectured that λ-definability exactly captured the informal concept of effective calculability (Church's Thesis).[17] The high point of this definability work was Church's negative solution of Hilbert's long-standing *Entscheidungsproblem* for first-order logic: first he proved in [Church, 1936b] that the convertibility problem for pure λ-calculus was recursively undecidable, then he deduced that no recursive decision procedure existed for validity in first-order predicate logic, see [Church, 1936a].

But this result was also proved independently almost immediately afterwards by Alan Turing in England. In Cambridge in the 1930s the topologist Max Newman had maintained a side-interest in the foundations, and in early 1935 he taught a course on this topic which Turing attended. Later that year Turing wrote his now famous paper on computability, [Turing, 1936]. He then visited Princeton and took a doctorate under Church (in 1936–38).

Incidentally, in the course of his doctoral work Turing gave the first published fixed-point combinator, [Turing, 1937b, term Θ]. This was seen as only having minor interest at that time, but in view of the later importance given to such combinators (see §8.1.2 below), we digress here to discuss them.

A *fixed-point combinator* is any closed term $\mathbf{Y}$ such that $\mathbf{Y}x$ converts to $x(\mathbf{Y}x)$.

[16] For an introduction to reduction-theory in general, see [Klop, 1992] or (less general) [Barendregt, 1981, §3.1], or (in λ) [Hindley and Seldin, 1986, §1C & Appendices 1 & 2] or [Barendregt, 1981, Chs. 11–15]. Some notation: a reducibility relation $\triangleright$ is *confluent*, or has the *Church-Rosser property*, iff $(U \triangleright X \ \&\ U \triangleright Y) \Longrightarrow (\exists Z)\ (X \triangleright Z \ \&\ Y \triangleright Z)$. It is *locally confluent* iff $(U \triangleright_1 X \ \&\ U \triangleright_1 Y) \Longrightarrow (\exists Z)\ (X \triangleright Z \ \&\ Y \triangleright Z)$, where "$\triangleright_1$" means one-step reducibility. A one-step reduction is called a *contraction*. A contraction $X \triangleright_1 Y$ is made by replacing an occurrence in X of a term called a *redex* by another term called its *contractum*. In λ, a (β-) redex is any term $(\lambda x.M)N$ and its contractum is $[N/x]M$. A term containing no redexes is called a *normal form*; it cannot be reduced.

[17] Comments on the history of Church's Thesis are in: [Kleene, 1981], [Kleene, 1952, §62], [Rosser, 1984], [Davis, 1982], [Gandy, 1988] and [Sieg, 1997].

Turing's was

$$(\lambda xy.y(xxy))(\lambda xy.y(xxy)).$$

The next one to appear was in [Rosenbloom, 1950, pp.130–131, Exs. 3e, 5f]. It was $\lambda x.\,\mathbf{W}(\mathbf{B}x)(\mathbf{W}(\mathbf{B}x))$, which is convertible to $\lambda x.\,(\lambda y.x(yy))(\lambda y.x(yy))$. The latter has often been called Curry's $\mathbf{Y}$; in fact Curry gave no explicit fixed-point combinator before [Curry and Feys, 1958, §5G], but the accounts of Russell's paradox in [Church, 1932, p.347] and [Curry, 1934a, §5] both mentioned $(\lambda y.N(yy))(\lambda y.N(yy))$, where N represented negation, and Curry's use of the latter term dated back to a letter to Hilbert in 1929.[18] Note that although fixed-point combinators are often used in λ-defining recursion, they are not really necessary for recursion on the natural numbers; none was used in the λ-representation of the recursive functions in [Kleene, 1936].

Returning to the Princeton group: their successes in pure λ-calculus were not really matched on the applied side, but the results were by no means completely negative. After the discovery of the Kleene-Rosser inconsistency, Church replaced his 1933 system by a free-variable logic based on the $\lambda\delta$-calculus, which he proved consistent by extending the Church-Rosser theorem. (In essence δ was a discriminator between normal forms; see [Church, 1935] or [Church, 1941, §20], and the review [Curry, 1937].) But the new system was too weak for much mathematics and seems to have played no role in Church's future work. In fact in the later 1930s Church retreated from the task of designing a general type-free logic to the much less ambitious one of re-formulating simple type theory on a λ-calculus base. In this he succeeded, publishing a smooth and natural system in [Church, 1940] that has been the foundation for much type-theoretic work since (see §5.1 and §8 below).

4.2 CL in the 1930s

The work of Curry through the 1930s continued his earlier study of the most basic properties of abstraction, universal quantification and implication in as general a setting as possible. He analysed the Kleene-Rosser proof of the inconsistency of Church's system, and by 1942 he had found in a logical system of his own a very simple inconsistency now called the *Curry paradox* [Curry, 1942b]. This showed combinatory completeness to be incompatible with unrestricted use of certain simple properties of implication in a very general setting.[19]

[18] Stated in [Curry and Feys, 1958, §5S]. By the way, Turing's fixed-point combinator satisfied $\mathbf{Y}x \rhd x(\mathbf{Y}x)$, but Rosenbloom's and Curry's only satisfied "=" not "$\rhd$". For more on fixed-point combinators, see [Barendregt, 1981, §§6.1, 6.5], [Hindley and Seldin, 1986, §3B], or [Curry *et al.*, 1972, §11F7].

[19] A general form of Curry's paradox runs thus, cf. [Hindley and Seldin, 1986, pp.268–269]. Suppose a system $\mathcal{S}$ of logic has a class $\mathcal{P}$ of expressions such that $X, Y \in \mathcal{P} \Longrightarrow (X \supset Y) \in \mathcal{P}$, and has an axiom-scheme $(X \supset (X \supset Y)) \supset (X \supset Y)$ (for all X, $Y \in \mathcal{P}$), and rules (Eq): X, $X = Y \vdash Y$ and (MP): X, $X \supset Y \vdash Y$.

For every $Z \in \mathcal{P}$, let Z' be $\lambda y.((yy) \supset ((yy) \supset Z))$, where y does not occur in Z, and let Z^* be $Z'Z'$. It is easy to prove that $Z^* = Z^* \supset (Z^* \supset Z) = (Z^* \supset (Z^* \supset Z)) \supset (Z^* \supset Z)$.

Curry also began to develop a *type theory*. A note on types dated 1928 occurs in his files, and he gave a talk on this topic in 1930 and wrote it up in [Curry, 1934a] and [Curry, 1936].[20] He took a different approach from Russell and from the one that Church later took: he added to his combinator-based logic a *functionality constant* $\mathbf{F}$, with the intention that an expression such as $\mathbf{F}abf$ should mean approximately $(\forall x)(x \in a \supset fx \in b)$.[21] This approach differed from Church's in that Curry's $\mathbf{F}abf$ did not imply that the domain of f was exactly a, only that it included a. In Curry's functionality-theory a term could have an infinite number of types. Indeed his axioms for $\mathbf{F}$ included axioms assigning an infinity of type-expressions to each atomic combinator $\mathbf{B}$, $\mathbf{C}$, $\mathbf{K}$, $\mathbf{W}$; for example (from [Curry, 1934a, p.586] or [Curry, 1936, p.378])

$$\text{Ax. (FB):} \quad (\forall x, y, z)\ \Big(\big(\mathbf{F}(\mathbf{F}xy)(\mathbf{F}(\mathbf{F}zx)(\mathbf{F}zy))\big)\ \mathbf{B}\Big).$$

Curry's approach was what we would now call *polymorphic*. In contrast, in a Church-style type-theory, for a term $f^{\alpha\to\beta}$, its domain would be exactly α and its type would be uniquely $\alpha \to \beta$.

Curry's use of types differed from Russell and Church in another way. Russell had introduced type-expressions to help him restrict his language to avoid certain troublesome expressions, and Church later had the same aim. But Curry allowed into his language all combinators, troublesome or placid, and simply assigned labels to some of them to tell us how they would behave.

Another early user of a polymorphic approach to types was Quine. In [Quine, 1937, pp. 78–79] he proposed restricting the comprehension axiom-scheme in set theory to formulae he called *stratified*, but which we would now call *typable*, and on p.79 he made a very clear argument in favour of a polymorphic approach. His proposed system is nowadays called NF for "New Foundations". Its history is beyond our scope, but the equivalence between stratification and typability was proved in [Newman, 1943, Thm. 4], and a general study of NF is in [Forster, 1995].

The *propositions-as-types* correspondence was noticed by Curry in the 1930s. In its simplest form this is an isomorphism between the type-expressions assignable to combinators and the implicational formulas provable in intuitionistic logic. The first hint of this, at least for atomic combinators, occurs in a note dated 1930 in Curry's files, and appeared in print in his papers [Curry, 1934a] and [Curry, 1934b].[22] In [Curry, 1934a, p.588] and [Curry, 1934b, p.850] he gave his axioms for implication the names of combinators, for example

$$\text{Ax. (PB):} \quad (\forall x, y, z)\ \big(\mathbf{P}(\mathbf{P}xy)(\mathbf{P}(\mathbf{P}zx)(\mathbf{P}zy))\big),$$

Suppose $Z \in \mathcal{P} \Longrightarrow Z^* \in \mathcal{P}$, for all Z. Then every Z in $\mathcal{P}$ is provable in $\mathcal{S}$. Because, the axiom-scheme gives $\vdash (Z^* \supset (Z^* \supset Z)) \supset (Z^* \supset Z)$, so by rule (Eq) twice, $\vdash Z^* \supset (Z^* \supset Z)$ and $\vdash Z^*$; hence by (MP) twice, $\vdash Z$. (Curry's treatment of the Russell paradox in [Curry, 1934a, pp.588–589] was interestingly similar to this argument.)

[20] We are very grateful to J. P. Seldin for showing us Curry's 1928 note (T 28.12.13.A, reproduced in [Seldin, 2002]).

[21] But only approximately; see the discussions of Π and $\mathbf{F}$ in [Seldin, 2008].

[22] We thank J. P. Seldin for a copy of Curry's 1930 note (T 30.07.15.B, reproduced in [Seldin, 2002]). The paper [Curry, 1934b] was based on a talk given in 1932.

where $\mathbf{P}xy$ was his notation for $x \supset y$. His name "(PB)" was clearly motivated by the axiom's similarity to Ax. (FB), although he did not state the **P**-to-**F** translation explicitly in print until later, [Curry, 1942a, p.60, footnote 28]. In [Curry, 1936, pp.391–394] he proved that if **F** was defined as

$$\mathbf{F} \quad \equiv \quad [a, b, f].\ (\forall x)(ax \supset b(fx))$$

in a system with combinators and suitable rules for $\forall$, then each **F**-axiom would be deducible from the corresponding **P**-axiom.

For non-atomic combinators the propositions-as-types correspondence was not mentioned by Curry until many years later; in fact the first discussion that included composite combinators was in [Curry and Feys, 1958, §9E]. There it was proved explicitly that when **F** was replaced by **P**, the set of all type-expressions assignable to the combinators became exactly the set of all formulas provable in intuitionistic implicational logic, provided type-assignment was made in a certain restricted system. This system had no $\forall$, had axiom-schemes instead of the previous **F**-axioms for the atomic combinators, and its main rule was:

$$\text{Rule } \mathbf{F}: \qquad \mathbf{F}ABF,\ \ AX \ \vdash\ B(FX)$$

(which would become the rule of *modus ponens* if **F** was replaced by **P** and F, X, FX were deleted). It had only one other rule, to ensure that interconvertible terms received the same types, [Curry and Feys, 1958, §9A3, Rule Eq′]; this rule did not affect the total set of assignable types, and was dropped in later studies of propositions-as-types. After dropping that rule, it became clear that there was also a correspondence between type-assignment deductions (or typed terms) and propositional deductions. This fact, and its analogue for λ, and some possible extensions of the correspondence, were pointed out by William Howard in a manuscript circulated in 1969 and published in [Howard, 1980]. The correspondence is often called *formulae-as-types*, following [Howard, 1980][23], or the *Curry-Howard isomorphism*.

The propositions-as-types correspondence was also discovered independently by several other workers in the 1960s; this period will be discussed further in §8.1.4 below.

Partly as a consequence of propositions-as-types, the authors of [Curry and Feys, 1958] were led to apply Gentzen's methods to type-theory, an idea which is standard now but was new then. This gave them the first published proof of the normalization theorem for typable terms, see §5.4 below.

In the 1930s another logician who made use of combinators was Frederic Fitch in Yale University. Fitch's work began with his doctoral thesis [Fitch, 1936], which described a type-free combinator-based logic that avoided Russell's paradox by

[23] Although this name was later criticised by Howard himself who said "a type should be regarded as an abstract object whereas a formula is the name of a type", [Howard, 1980, p.479].

By the way, in [Frege, 1879, §13] the axioms for implication were the principal types of **K**, **S** and **C**. These might be seen, via propositions-as-types, as a partial anticipation of the concept of combinator basis, at least for the typable terms. But this is probably far-fetched; in particular, [Frege, 1879] had no explicit concept of completeness.

forbidding certain cases of abstraction $[x_1, ..., x_n].X$ where X contained repeated variables. He then developed over several decades a system CΔ, very different from the one in his thesis but still type-free and combinator-based, with unrestricted quantification and abstraction; see [Fitch, 1963]. The set of theorems of CΔ was not recursively enumerable, [Fitch, 1963, p.87], but the system was adequate for a considerable part of mathematical analysis and had a (non-finitary) consistency proof. He presented a further modified system in his student-level textbook on CL, [Fitch, 1974].

Incidentally, combinators are not the only way of eliminating bound variables: an alternative technique is *algebraic logic*. Very roughly speaking, while CL algebraizes abstraction the latter algebraizes quantification. In 1936 Quine sketched how this could be done in principle by means of an algebra of relations, [Quine, 1936b]. (Such algebras had been published earlier in [Peirce, 1883] and [Schröder, 1905, Vol.III].) In 1938 Leopold Löwenheim proposed to Alfred Tarski that all of mathematics could be expressed in a Peirce-Schröder algebra, see [Löwenheim, 1940], and in 1941 Tarski developed an abstract theory of binary relations, called the theory of *relation algebras*, and used it to prove that set theory could be formalized without variables, see [Tarski, 1941] and [Tarski, 1953]. A detailed formalization of set theory without variables was carried out in the book [Tarski and Givant, 1987], completed just before Tarski died in 1983.

However, for a formalization of first-order logic in general, Tarski's theory was inadequate, and to remedy this he and his co-workers developed the theory of *cylindric algebras* from the late 1940s onward. A similar algebraic treatment of bound variables was also given by the *polyadic algebras* of Paul Halmos in 1956, see [Halmos, 1962]. Other significant contributions to the algebraic approach were made by [Bernays, 1958], and [Craig, 1974]. William Craig had previously contributed to CL, see [Curry and Feys, 1958, §§5H, 7E].

For more on the history of algebraic logic, see the introductions to [Tarski and Givant, 1987] and [Givant and Andréka, 2002]. A good comparison of algebraic logic with type-free CL is in [Quine, 1972]. From the viewpoint of a semantics based on a standard set theory such as ZF or NBG, an advantage of algebraic logic is that it retains the restrictions of first-order logic, and therefore its semantic simplicity; as noted earlier, unrestricted type-free combinators are much harder to interpret in standard set theories.

A different algebraic approach was taken by the mathematician Karl Menger. Menger's method involved an algebraization of a limited form of function-composition, and was expounded by him in a series of papers from the 1940s to the '60s, for example [Menger, 1944] and [Menger, 1964]. But his system is more restricted than Tarski's, and less mathematics has been developed in it; see the review [Lercher, 1966].

5 1940s AND 1950s: CONSOLIDATION

For many years after the mid-1930s, neither λ nor CL attracted much interest among logicians in general. Both systems would later play a natural role in studies of higher-order logic, but the formal metatheory of that topic had hardly begun to be developed at that early period. The λ-formalism had been designed to describe substitution and application in detail and seemed off-puttingly heavy for other purposes. It had given the first proof of the unsolvability of the *Entscheidungsproblem*, but Turing's proof via his machine model of computation was much more transparent, and even Kleene preferred other systems for his subsequent development of computability theory.

However, on the expository side, an advance in opening up λ to non-specialists was made by Church with his readable introductory textbook [Church, 1941]. And Paul Rosenbloom included a section on CL in his general logic textbook [Rosenbloom, 1950, Ch.III, §4]. Advances were made in both subjects on the technical side too, and these helped to lay a foundation for the expansion that came later. They can be split roughly into four main themes: simple type theory, abstract reduction theory, reductions in CL and λ, and illative systems.

5.1 Simple type theory

Church's simple type theory was a function-based system, stemming from ideas of Frank Ramsey and Leon Chwistek in the 1920s, for simplifying the type theory of [Russell and Whitehead, 1913].[24] Church lectured on his system in Princeton in 1937–38 before publishing it in [Church, 1940], and his lectures were attended by Turing, who later made some technical contributions. These included the first proof of the system's weak normalization theorem, although this lay unpublished for nearly 40 years, see [Turing, 1980] and §8.2 below.

Church's system was analysed and extended in a series of Princeton Ph.D. theses from the 1940s onward, of which perhaps the best known are Leon Henkin's in 1947, published in [Henkin, 1950], and Peter Andrews', published in [Andrews, 1965]. Henkin gave two definitions of model of typed λ (*standard* and *general* models),[25] and proved the completeness of simple type theory with respect to general models. Andrews extended Church's system to make a smooth theory of tranfinite types.

Church's own interests underwent some evolution after his 1941 textbook appeared. Much of his time and energy was spent on the business of the Association for Symbolic Logic, of which he had been a co-founder in 1936 and in which he was still a driving force, and on editing the Journal of Symbolic Logic, especially its very comprehensive Reviews and Bibliography sections. On the research side he became interested in intensional logic and the logic of relevant implication.

[24] See [Gandy, 1977] for a sketch of the origins of Church's system (and of Turing's contributions).

[25] See §8.1.2 below for the difference between these; details are in [Henkin, 1950, pp.83–85].

For the former he made axiomatizations in [Church, 1951a] and [Church, 1973], both based on modifications of his type theory. For relevant implication he gave a formulation, in [Church, 1951b], which was essentially equivalent, under the propositions-as-types correspondence, to the simple type theory of pure λI-terms.

5.2 Abstract reduction theory

Probably the first publication on abstract reduction-theory was a paper by the mathematician Axel Thue on rewriting finite trees [Thue, 1910].[26] Thue's paper even included a lemma that local confluence implies confluence if every contraction strictly decreases a finite measure [Thue, 1910, §4 Satz 3], very like Newman's lemma of 30 years later that local confluence implies confluence if all reduction paths are finite [Newman, 1942, Thm. 3].

But the main inspiration for abstract work in this field was [Newman, 1942]. In the 1920s, Newman proved a crucial confluence conjecture in combinatorial topology, and a decade later his interest in logic (and probably his link with Turing) led him to study the confluence theorem for λ and the method of residuals that Church and Rosser had used in its proof [Church and Rosser, 1936].[27] The first part of [Newman, 1942] was a study of confluence in a very general setting; in the second, Newman gave a set of abstract axioms about the concept of residual, from which he deduced confluence. Unfortunately, due to an error involving variables, his axioms failed to cover λ as a special case. This was first corrected about 20 years later by David Schroer in his Ph.D. thesis [Schroer, 1965].

Newman's analysis led in the 1970s to general methods for proving confluence and other key theorems, such as standardization, for systems of λ augmented by extra operators, for example a primitive recursion operator. (A summary of some results is in [Hindley, 1978b, pp.269–270].) Three influential abstract studies of confluence were [Rosen, 1973], [O'Donnell, 1977] and [Huet, 1980].

In the 1980s the abstract theory of reductions was re-organised completely and developed by Jan-Willem Klop, beginning with his penetrating thesis [Klop, 1980], and since then the important abstract theories of *combinatory reduction systems* and *term rewriting systems* have evolved. These are being actively developed today, but go far beyond the boundaries of λ and CL and are the subjects of good surveys in [Klop, 1992] and [Klop *et al.*, 1993] and a comprehensive textbook [Terese, 2003], so we shall not give details of them here.

[26] See [Steinby and Thomas, 2000], which contains an English summary of [Thue, 1910].

[27] If a term X contains redex-occurrences $R_1, ..., R_n$, and $X \triangleright_1 Y$ by contracting R_1, certain parts of Y can be considered to be descendants of $R_2, ..., R_n$ in a sense; they are called *residuals* of $R_2, ..., R_n$; for details in λ, see [Curry and Feys, 1958, §§4B1–2] or [Hindley and Seldin, 1986, Def. A1.3], or [Church and Rosser, 1936, pp.473–475]. If we reduce Y by contracting one of these residuals, and then a residual of these residuals, etc., the reduction $X \triangleright_1 Y \triangleright$ ... is called a *development* of $R_1, ..., R_n$. If it is finite and its last term contains no residuals of $R_1, ..., R_n$, it is called *complete*.

5.3 Reductions in CL and λ

Besides abstract reduction theory, there was also progress in analysing the two most important concrete reductions, weak reduction in CL and β-reduction in λ. As mentioned earlier, Church and Rosser had begun the study of these in the paper containing their confluence proof, [Church and Rosser, 1936].

After 1950 their programme was continued by Curry. His research had been suspended completely during the Second World War, like that of many other academics, and did not recover momentum until quite late in the 1940s. Then for about 20 years he put his main effort into compiling two large monographs containing much new work, [Curry and Feys, 1958] and [Curry *et al.*, 1972]. Part of this effort consisted of a detailed study of $\lambda\beta$-reduction, including the first proof of its standardization theorem, [Curry and Feys, 1958, §4E1]. He also made the first study of $\lambda\beta\eta$-reduction, including its confluence and η-postponement theorems, [Curry and Feys, 1958, §4D]. (But his proof of the latter contained a gap, which was noticed and remedied fifteen years later by Robert Nederpelt, [Nederpelt, 1973, Thm. 7.28].)

In 1965 the thesis of Schroer gave the first statement and proof of the finiteness-of-developments theorem for $\lambda\beta$, [Schroer, 1965, Part I, Thm. 6.20]. But Schroer's work was never published and this theorem was not re-discovered until about 8 years later, see §7.2 below.

Weak reduction in CL is much simpler than β-reduction in λ, and was given a straightforward systematic analysis in [Curry *et al.*, 1972, §11B].

Curry also defined a *strong reduction* for CL analogous to $\lambda\beta\eta$-reduction, [Curry and Feys, 1958, §6F].[28] His purpose was to make later discussions of logic applicable equally to languages based on CL and on λ. Strong reduction indeed allowed this to be done to a significant extent, but its metatheory turned out to be complicated, and although simplifications were made by Roger Hindley and Bruce Lercher in the 1960s which produced some improvement, it eventually seemed that wherever strong reduction could be used, it would probably be easier to use λ instead of CL. Strong reduction was therefore more or less abandoned after 1972. However, in the 1980s a β-strong reduction was defined and studied by Mohamed Mezghiche, see [Mezghiche, 1984], and the idea of working with a combinatory reduction which is preserved by abstraction but avoids the complexity of bound variables is still tantalisingly attractive.

5.4 Illative systems

Turning from pure to applied CL and λ: although type-theory made important advances in the 1950s, some logicians felt that type-restrictions were stronger than necessary, and that type-free higher-order logic was still worth further study. Most

[28] In CL, $\beta\eta$-strong reduction $\succ_{\beta\eta}$ is defined by adding to the rules defining weak reduction the extra rule (ξ): *if* $X \succ_{\beta\eta} Y$ *and* x *is any variable, then* $[x]_{abcf}.X \succ_{\beta\eta} [x]_{abcf}.Y$, where $[x]_{abcf}$ is evaluated by abstraction-algorithm (abcf), see §3 above.

systems of such logic contain analogues of combinators in some form, but here we shall only mention those in which CL or λ were more prominent.

Curry's work on applied systems (which he called *illative*, from the Latin "illatum" for "inferred") was published in the books [Curry and Feys, 1958] and [Curry *et al.*, 1972]. He never gave up his interest in type-free logic, although one of his main themes of study at this time was the relation between types and terms.

In his 1958 book with Robert Feys, perhaps the authors' most important contribution in the illative sections was to introduce Gentzen's techniques into type theory. As we mentioned in §4.2, this approach was not as obvious in 1958 as it is now. Feys had written on combinatory logic and on Natural Deduction in [Feys, 1946a; Feys, 1946b], and Curry had emphasized the importance of Gentzen's methods in a series of lectures on first-order logic at the University of Notre Dame in 1948. In [Curry and Feys, 1958, §9F] they presented a sequent-based system of type-assignment and a proof of a cut-elimination theorem for this system, from which they deduced the weak normalization theorem for simple type theory [Curry and Feys, 1958, §9F6, Cor. 9F9.2]; this was the first published proof of this theorem. (Other proofs will be discussed in §8.2.) Further type-theoretic properties described in that book included the propositions-as-types correspondence in [Curry and Feys, 1958, §9E], see §4.2 above and §8.1.4 below, and the link with grammatical categories in [Curry and Feys, 1958, §8S2], see §8.1.1 below.

In [Curry *et al.*, 1972], Curry's main aim was to make a thorough analysis of basic logical concepts in the settings of λ and CL, with as few restrictions as possible and an emphasis on constructive methods.[29] The book contained a detailed comparison of logical systems based on three different concepts: (i) $\mathbf{F}$, the functionality constant described in §4.2 above, (ii) Ξ, where ΞAB was approximately equivalent to $(\forall x)(Ax \supset Bx)$, and (iii) $\forall$ and $\supset$. The formulation of these systems benefitted greatly from results and simplifications introduced by Jonathan Seldin in his thesis [Seldin, 1968], but many variants of each system were discussed and the resulting treatment was rather complex.

A proposal for a generalized functionality concept $\mathbf{G}$ was made by Curry in [Curry and Feys, 1958, Appendix A] and [Curry *et al.*, 1972, §§15A1, 15A8]. In terms of $\forall$ and $\supset$, $\mathbf{G}ABF$ was intended to mean approximately $(\forall x)(\, Ax \supset (Bx)(Fx)\,)$.[30] This proposal was developed into a fully-fledged higher-order type-theory by Seldin in 1975, see [Seldin, 1979] or [Hindley and Seldin, 1986, 1st edn., Ch. 17]. Seldin later became interested in Thierry Coquand's calculus of constructions, which was stronger, see [Seldin, 1997], and thus his work on Curry-style illative systems gradually merged with the work of others on higher-order systems. These will be discussed in §8.3 below.

Returning to the type-free approach to higher-order logic: in the 1950s the main builder of type-free systems was Fitch. As described in §4.2, his work on this

[29] For further comments on Curry's programme see [Seldin, 1980a, pp.22–27].

[30] In a $\mathbf{G}$-based system the type of the output-value of a function may depend on the value of its argument. Type systems with this property are called *dependent*. Other systems of dependent types are mentioned in §8.3.

continued into the 1970s. Some of his technical ideas were appreciated and used by other workers, for example Corrado Böhm, see §6.3, and Solomon Feferman, see [Feferman, 1984, p.106]. But his formal systems themselves seem to have been little used.

The system-building tradition was continued by others, however. In 1974 in his thesis for Doctorat d' Etat, André Chauvin produced a thoughtful and interesting analysis of type-free higher-order logic and the role of combinators, and described a formal system based on the notion of *partial* operation or function. Unfortunately only a little of this thesis was published, [Chauvin, 1979].

The concept of partial function was also behind several substantial papers by Feferman from 1975 through the 1980s, in which he examined the possibilities for type-free systems, particularly as foundations for constructive mathematics and for category theory, and proposed several actual systems in which "partial" combinators played a role, see for example [Feferman, 1975a; Feferman, 1975b; Feferman, 1977; Feferman, 1984] and [Aczel and Feferman, 1980].

An analysis of the foundations based closely on λ was made by Peter Aczel in an influential paper [Aczel, 1980].

From 1967 onward, work on type-free illative systems was undertaken by Martin Bunder. He made a careful study of the inconsistencies that had arisen in past systems, and proposed several new systems that avoided these, including some in which all of ZF set theory can be deduced; see, for example, [Bunder, 1983a] and [Bunder, 1983c], depending on [Bunder, 1983b]. A good short overview is in [Barendregt *et al.*, 1993].

With Wil Dekkers and Herman Geuvers, Bunder and Barendregt published useful comparisons between their type-free illative systems and other higher-order formal systems such as Pure Type Systems: [Dekkers *et al.*, 1998] and [Bunder and Dekkers, 2005].

In 1990, Randall Holmes formulated systems of λ and CL in which abstraction was restricted by stratification rules inspired by the stratification concept in Quine's "New Foundations" system [Quine, 1937]: see [Holmes, 1991], or [Holmes, 1995] for a short overview.

In Moscow, independently of all the above, a series of type-free combinator-based systems was proposed by Alexander Kuzichev from the 1970s onward; see, for example, [Kuzichev, 1980; Kuzichev, 1983; Kuzichev, 1999] and the references therein.

6 PROGRAMMING LANGUAGES

Returning to the 1960s: at about this time, λ and CL began to attract the interest of a new group outside the community of logicians, namely computer scientists involved in the theory and practice of programming languages. Perhaps the first published suggestion that CL or λ be used directly as a programming language came from the logician Fitch in 1957, although for only a restricted class of problems: in his paper [Fitch, 1958] he showed how sequential logic-circuits could be

represented in CL.[31] But the main characters concerned were computer scientists: we shall look at three of these.

6.1 John McCarthy and LISP

From 1956 to '60, John McCarthy in the U.S.A. developed the computer language LISP, which was a list-processing language with a function-abstraction facility. McCarthy's goal was to apply LISP eventually to problems in non-numerical computation, especially in the newborn field of *artificial intelligence*, and to encourage the style of program-organization that is nowadays called *functional programming*.[32]

LISP was not directly based on λ-calculus, although it owed something to λ (and its abstraction notation was even called "LAMBDA"). LISP's substitution procedure was not defined exactly the same as in λ, in particular identifiers were handled according to so-called "dynamic binding" rules. This choice simplified LISP's implementation by means of interpreters, but it greatly complicated the use of bound variables by the programmer.[33]

McCarthy's emphasis on a functional approach to programming also had influence on the theoretical side, where his contributions included two important papers on a mathematical basis for the theory of computation, [McCarthy, 1963b; McCarthy, 1963a]. There we find a systematic use of conditional expressions within a functional formalism for recursively defining functions over symbolic expressions like the S-expressions of LISP, together with a thorough study of their formal properties. Equalities among recursive functions expressed in that formalism were proved by means of *recursion induction*, to be mentioned again in §8.1.2 in connection with its generalization as fixed-point induction by Scott [Scott, 1969e]. Along with recursive definitions of functions, [McCarthy, 1963a, §2.6] also discussed recursive definitions of *sets*, like the definition of the set of S-expressions itself as a solution of the equation $S \cong A + (S \times S)$.

6.2 Peter Landin

In the early 1960s in England, Peter Landin proposed the use of λ-terms to code constructs of the programming language Algol 60, see [Landin, 1965].[34] While for LISP a precise correspondence with λ was hindered by the use of dynamic binding, for Algol its block structure perfectly matched the way names were handled in λ. In fact, Landin's work made it possible to look at λ itself as a programming language,

[31] We have recently learned that actually Fitch was not the first; Curry suggested using combinators in programming, in a talk in 1952, [Curry, 1954]; see [Seldin, 2008].

[32] This name dates back at least to a 1963 tutorial on LISP by P. M. Woodward [Fox, 1966, p.41].

[33] See [Moses, 1970] for comments. For the ideas behind LISP see [McCarthy, 1960], [McCarthy, 1963a]; for its history to 1996 see [McCarthy, 1981], [Steele and Gabriel, 1996]. One notable descendant of LISP with the variable-binding problem fixed is Scheme, see [Dybyg, 1996], dating from about 1975.

[34] More precisely, λ-terms extended with some primitive operations for dealing with jumps.

and one especially suited for theoretical purposes. Taking this standpoint, and in parallel with his translation of Algol, Landin described in 1963 an abstract machine for reducing λ-terms considered as programs, the *SECD-machine* of [Landin, 1964, pp. 316–318] and [Landin, 1966a]. This consisted of a transition system whose states were made of four components:

- a *stack* S for storing intermediate results of evaluation,
- an *environment* E, that associates values to free identifiers,
- a *control* C consisting of a list of expressions that drives the evaluation process,
- a *dump* D that records a complete state of the machine, of the form (S', E', C', D').

The transition rules of the machine implemented the order of evaluation now known as *call by value*, so their results did not agree exactly with those of leftmost-first reduction of λ-terms. While the design of the SECD-machine was influenced by similar earlier devices for the interpretation of programming languages, notably the LISP interpreter devised in [Gilmore, 1963],[35] it was however the first to be presented in a formal way, and may fairly be said to have opened the way to the study of the *formal* operational semantics of λ-calculus considered as a programming language, that would find a systematic exposition a decade later starting with [Plotkin, 1975].

6.3 Corrado Böhm: λ-calculus as a programming language

In the above developments λ and CL began to contribute ideas to the theory of programming languages, and their influence is still active; but in the opposite direction, programming experience had only a limited impact on the "pure" theory of λ and CL at that time. The exception was the work of Corrado Böhm from the early 1960s onward.[36]

Böhm gained his Ph.D. in 1951 at the ETH in Zurich, with a thesis which included the first-ever description of a complete compiler in its own language, [Böhm, 1954]. While working on his thesis he came in contact with Bernays, who introduced him to the classical formalisms for computable functions. These included Turing machines and Post and Thue systems, but the variety of these models of computation led Böhm in the following years to search for an inclusive formalism. This aim was the origin of his interest in λ-calculus, with which he became acquainted around 1954 from Church's textbook [Church, 1941], but whose importance he realized only at the end of the 1950s after noticing that functional abstraction was used in LISP. A colleague, Wolf Gross, drew his attention to [Curry and Feys, 1958] and to the third chapter of [Rosenbloom, 1950], and in the years

[35] See [Landin, 1966a, §20] for a discussion of related work. A textbook treatment of the SECD-machine is contained in [Burge, 1978].

[36] Our reconstruction of Böhm's work has been helped by the scientific biography in [Dezani *et al.*, 1993, pp. 1–8], and a personal communication by Böhm, April 11th, 2001.

1960-61 Böhm and Gross together began to study combinators, λ-calculus and recursive function theory. By 1964 they had realized that λ-terms and combinators could be used directly as the basis of "a very flexible description language for machines, programs and algorithms", [Böhm and Gross, 1966, p.36]. This was the starting point of the CUCH language (a system using both CUrry's combinators and CHurch's λ-terms), which was developed independently of the ideas that were emerging at about the same time from the work of Strachey and Landin, but was influenced partly by the use of λ-abstraction notation in [McCarthy, 1963a, §3] (and by the use of β-reduction in Paul Gilmore's interpreter [Gilmore, 1963, §3]), and especially by the use of combinators to describe sequential circuits in [Fitch, 1958]. The latter gave in fact one of the first applications of CUCH (described in [Böhm and Giovannucci, 1964] together with the representation of analogue circuits), as well as the encoding of the operations of Iverson's programming language APL in unpublished work of Marisa Venturini Zilli in 1962-64.[37]

An abstract CUCH-machine was described in [Böhm and Dezani, 1972], and had an actual implementation in the 1990s, in unpublished work by Stefano Guerrini. The CUCH machine is of interest for the history of abstract interpreters for the λ-calculus, for it used a leftmost-first reduction strategy instead of the call-by-value order realized by the more widely known SECD-machine of Landin.

But the main influence of Böhm and his students on λ and CL was through discoveries about the pure systems. The presentation of CUCH in [Böhm, 1966] and [Böhm and Gross, 1966] set up a uniform language for formulating technical problems about the syntax of untyped λ and CL, and the solution of such problems dominated much of Böhm's later research activity. We shall discuss these in §7.1.

7 SYNTACTICAL DEVELOPMENTS

The revival of interest in λ and CL that began in the 1960s brought with it some new advances in the study of their syntax.

Work by proof-theorists on Gödel's *Dialectica* paper [Gödel, 1958], stimulated particularly by Georg Kreisel in Stanford University, led to normalization proofs for various typed extensions of λ and CL from 1963 onward. These will be discussed in §8.2 below.

In 1963 Dana Scott gave lectures on λ to an autumn seminar on foundations at Stanford [Scott, 1963]. These were partly motivated by "a certain lack of satisfaction with systems for functions suggested in papers of John McCarthy", with the hope "that the underlying philosophy will ... lead to the development of a general programming language for use on computers" [Scott, 1963, p. 1.2]. Scott's syntactical innovations included a new representation of the natural numbers for which the predecessor combinator was much simpler than those known for Church's numerals,[38] and the following neat and general new undecidability theorem for λ: if

[37] Almost 20 years later, [Böhm, 1982] described a combinatory representation of Backus' systems of functional programming, [Backus, 1978], that are in many respects similar to APL.

[38] Scott's numerals were $\mathbf{0} \equiv \mathbf{K}$, $\mathbf{succ} \equiv \lambda uxy.\, yu$, see [Scott, 1963, §3] or [Curry *et al.*, 1972,

two non-empty sets of terms are closed under expansion (i.e. reversed reduction), then the pair cannot be recursively separated, i.e. there is no total recursive function which outputs 1 on the members of one set and 0 on the other; see [Scott, 1963, §9] or [Barendregt, 1981, §6.6] or [Curry *et al.*, 1972, §13B2]. Incidentally, Curry independently proved a similar undecidability theorem, though his was slightly less general than Scott's and was made a few years later (but before seeing Scott's theorem), see [Curry, 1969b, p.10] or [Curry *et al.*, 1972, p.251, footnote 7].

Scott's next major contribution to the theory of λ was on the semantic side in 1969 and radically changed both the subject itself and its more general context, the study of higher-order functions; see §9.1 below.

In 1971 appeared the influential thesis of Henk Barendregt, [Barendregt, 1971]. Part of its motivation was semantical, so we shall discuss it in §9.2 p.779, but its results and methods were syntactical and stimulated much work on syntactical properties of pure λ in the 1970s. Some of this work will be described in §7.2, and many of its pre-1981 results were incorporated into the monograph [Barendregt, 1981], which became the standard reference for the core properties of λ.

7.1 Contributions from the programming side

In Italy, the work of Böhm on the λ-based language CUCH led him to ask a series of technical questions about the syntax of λ which came to have an important influence on the development of the subject, both through the questions themselves and through the later activities of the students he attracted to his work. The first of these questions, at the end of a talk given in 1964 [Böhm, 1966, p.194], was whether it was possible to discriminate between any two distinct closed normal forms F, G by constructing a λ-term $\Delta_{F,G}$ such that

$$\Delta_{F,G}F =_{\beta\eta} \lambda xy.x, \qquad \Delta_{F,G}G =_{\beta\eta} \lambda xy.y.$$

A few years later the answer "yes" was proved for $\beta\eta$-normal forms by Böhm himself, [Böhm, 1968]; this result is now known as *Böhm's theorem*.[39] Böhm's proof of his theorem was analysed by Reiji Nakajima and by Barendregt independently, see [Nakajima, 1975] and [Barendregt, 1977, §6], and from these analyses came the now well-known concept of *Böhm tree*, as well as the use of this concept in the study of models of λ (see §9.2 below).[40] In [Barendregt, 1981, Ch. 10] there is

pp. 260–261]. For them, **pred** $\equiv \lambda x.\, x\mathbf{0I}$. Some predecessors known for the Church numerals are given in [Curry *et al.*, 1972, pp. 218, 226]. Other numeral systems have been proposed and used, for example in [Barendregt, 1976, §4.1], in [Wadsworth, 1980], and by Böhm, see below. Numeral systems in general have been discussed, with examples, in [Curry *et al.*, 1972, §§13A1, 13C1] and [Barendregt, 1981, §§6.4, 6.8].

[39] A well organised proof of Böhm's theorem is in [Krivine, 1990, pp. 68–74]. A fuller proof which has been mechanised in the language CAML is in [Huet, 1993]; it deals with problems about bound variables which are usually ignored but become serious when infinite trees are involved.

[40] Although for terms in normal form the Böhm tree concept had already appeared in [Böhm and Dezani, 1974].

a careful account of the ideas behind the proof (including the name *Böhming out* for the technique of extracting subtrees of a Böhm tree by means of contexts).

Another conjecture of Böhm's, made in [Böhm, 1968, p.4], was what is now called the *Range theorem*: the *range* of a combinator F is the set of all $\beta\eta$-convertibility classes of closed λ-terms of the form FM, and the theorem states that this set is either infinite or a singleton. This theorem was proved within a year by Myhill and Barendregt independently. A constructive version of the theorem, prompted by a remark of Dirk van Dalen, together with some historical remarks on the various proofs, is in [Barendregt, 1993].

Böhm's aim of using CUCH as a real programming language led him to study various ways of coding data structures as λ-terms. He coded natural numbers as the Church numerals [Böhm, 1966, Appendix II], although he also considered alternative representations, for example

$$\mathbf{n} \quad \equiv \quad \lambda x.\langle \underbrace{x, \ldots, x}_{n \text{ times}} \rangle$$

as described in [Böhm, 1966, Appendix VI]. Vectors $\langle X_1, \ldots, X_n \rangle$ were represented as terms of the form $\lambda x.xX_1 \ldots X_n$ with projections U_i^n defined by $\lambda x_1 \ldots x_n.x_i$, [Böhm, 1966, Appendix IV]. Böhm saw both data and programs as higher-order functions that should be represented by terms in normal form (or by strongly normalizing terms). This led him to the problem of structuring the set of all normal forms algebraically, which became especially important in his work. He wished to find an algebraic structure of normal forms suitable for representing data structures as initial algebras, and over which iterative, primitive recursive and even total recursive functions could be represented by terms in normal form. He explored this representation from 1983 onwards, and reported the results in a long series of works: [Böhm and Berarducci, 1985], [Böhm, 1986; Böhm, 1988b; Böhm, 1988a; Böhm, 1989]. The first of these, with Alessandro Berarducci in Rome, investigated the representation of data structures in second order λ-calculus and, as we shall discuss in §8.3, was very influential in investigations of polymorphic λ-calculi.

The quest for algebraic structure in combinatory logic led also to the study of the monoid structure of combinators (modulo η-conversion), extending an early idea of Church [Church, 1937], and even to combinatory groups through the study of invertible terms. In 1976 his co-worker and former student Mariangiola Dezani[41] in Turin made a characterization of the invertible normalizable terms of $\lambda\beta\eta$-calculus, [Dezani, 1976], and her result was extended to all terms by Klop and Jan Bergstra, [Bergstra and Klop, 1980]. With Francoise Ermine, Dezani then gave a description of maximal monoids of normal forms, using an early version of intersection types, [Dezani and Ermine, 1982]. Monoids of combinators were studied further in [Böhm and Dezani, 1989]. (A detailed account of the results in this area until 1980 is in [Barendregt, 1981, Ch. 21].)

[41] Married name Mariangiola Dezani Ciancaglini.

Taking a similar algebraic approach, around 1977 Böhm proposed a simple reinterpretation of combinatory logic by regarding combinators as generalized numerals [Böhm, 1979; Böhm, 1980]; this was suggested by the fact that exponentiation $\mathbf{m^n}$ of Church numerals $\mathbf{m}, \mathbf{n}$ is just $\mathbf{nm}$. He regarded combinators as a structure with sum, product and exponentiation with the usual arithmetical axioms, plus a non-arithmetical constant for pairing.

James H. Morris at the Massachusetts Institute of Technology was another influential contributor who, like Böhm, was motivated by an interest in programming languages; his thesis [Morris, 1968] acknowledged an intellectual debt to Landin, although most of the results in it concerned the syntax of untyped and simply typed λ-calculus.

That thesis contained the first formulation of "pure" simply typed λ in isolation without the extra logical axioms and rules that the systems of Church and Curry had contained. Morris' formulation was what is nowadays called "Curry-style"; types were assigned to untyped pure λ-terms. In [Morris, 1968, p.107] he proved weak normalization by a method essentially the same as in Turing's unpublished note, see §8.2 below, and in [Morris, 1968, pp.100–105] he described and verified a test for typability, see §8.5 below.

Morris' thesis also contained perhaps the first application of Böhm's Theorem, in a result which, although syntactical, later played a significant role in the model theory of λ. Morris suggested looking at a λ-term M as a partial function consisting of those pairs $\langle\langle M_1, \ldots, M_n\rangle, N\rangle$ such that reduction starting at $MM_1 \ldots M_n$ halts with the normal form N. The problem with this idea was that $\beta\eta$-conversion does not coincide with equality of partial functions, so Morris introduced the following preorder on terms [Morris, 1968, p.50]:

> Given λ-terms M, N: M *extends* N if, for all contexts $C[\ ]$, if $C[M]$ has a normal form then $C[N]$ has the same normal form.

The equivalence relation on terms generated by the extension preorder came later to be called *contextual equivalence*. Morris proved it to be, essentially, the largest congruence on (pure) λ-terms that non-trivially extends $\beta\eta$-convertibility, by applying the newly-published Böhm Theorem [Morris, 1968, p.53, Cor. 5].

A point worth noting in Morris' definition of the extension preorder was the use of *contexts*: these became an ubiquitous tool for the definition of equivalences on both untyped and typed terms (see §9.2 and §10 later). Morris also proved that the fixed-point combinator $\mathbf{Y}$ commonly attributed to Curry yields the least fixed point under the extension preorder [Morris, 1968, Ch. III, §C, Cor. 7(a)], by a complicated analysis of reductions. This result is used to justify, by way of example, a form of McCarthy's principle of Recursion Induction.

Near the end of his thesis, Morris argued for the usefulness of allowing circular type-expressions in simply typed λ-calculus, like the solution of the type equation $\alpha = \alpha \to \beta$, remarking that "we have only a limited intuition about what kind of an object a circular type expression might denote (possibly, a set of functions, some of which have themselves as arguments and values)", [Morris, 1968, p.123].

The answer to this puzzle was to be found less than one year later by Dana Scott with his set-theoretic model for the λ-calculus, see §9.1 below.

7.2 Theory of reductions

Ever since the original proof of the confluence of $\lambda\beta$-reduction in [Church and Rosser, 1936], a general feeling had persisted in the logic community that a shorter proof ought to exist. The work on abstract confluence proofs described in §5.2 did not help, as it was aimed mainly at generality, not at a short proof for $\lambda\beta$ in particular.

For CL, in contrast, the first confluence proof was accepted as reasonably simple; its key idea was to count the simultaneous contraction of a set of non-overlapping redexes as a single unit step, and confluence of sequences of these unit steps was easy to prove, [Rosser, 1935, p.144, Thm. T12].

Then in 1965 William Tait presented a short confluence proof for CL to a seminar on λ organized by Scott and McCarthy at Stanford. Its key was a very neat definition of a unit-step reduction by induction on term-structure. Tait's units were later seen to be essentially the same as Rosser's, but his inductive definition was much more direct. Further, it could be adapted to $\lambda\beta$. (This possibility was noted at the seminar in 1965, see [Tait, 2003, p.755 footnote]). Tait did not publish his method directly, but in the autumn of 1968 he showed his CL proof to Per Martin-Löf, who then adapted it to $\lambda\beta$ in the course of his work on type theory and included the $\lambda\beta$ proof in his manuscript [Martin-Löf, 1971b, pp.8–11, §2.5].

Martin-Löf's $\lambda\beta$-adaptation of Tait's proof was quickly appreciated by other workers in the subject, and appeared in [Barendregt, 1971, Appendix II], [Stenlund, 1972, Ch. 2] and [Hindley *et al.*, 1972, Appendix 1], as well as in a report by Martin-Löf himself, [Martin-Löf, 1972b, §2.4.3].[42]

In λ, each unit step defined by Tait's structural-induction method turned out to be a minimal-first development of a set of redexes (not necessarily disjoint). Curry had introduced such developments in [Curry and Feys, 1958, p.126], but had used them only indirectly; Hindley had used them extensively in his thesis, [Hindley, 1969a, p.547,"MCD"], but only in a very abstract setting. They are now usually called *parallel reductions*, following Masako Takahashi. In [Takahashi, 1989] the Tait-Martin-Löf proof was further refined, and the method of dividing reductions into these unit steps was also applied to simplify proofs of other main theorems on reductions in λ.

Tait's structural-induction method is now the standard way to prove confluence in λ and CL. However, some other proofs give extra insights into reductions that this method does not, see for example the analysis in [Barendregt, 1981, Chs. 3, 11–12].

[42]The earlier manuscript [Martin-Löf, 1971b] was not published, as the system described in it was inconsistent from the standpoint of the propositions-as-types correspondence, see §8.1.4 below; but its confluence proof for $\lambda\beta$ was not faulty.

Besides confluence, major themes in the study of $\lambda\beta$-reductions have been given by theorems on *finiteness of developments*, *standardisation* and the *equivalence of complete developments*, all depending either directly or indirectly on the concept of *residual*. The first two of these theorems were proved for λI-terms in [Church and Rosser, 1936, Lemma 1], and the second was proved for CL in [Rosser, 1935, p.142, Thm. T9]; but their analogues for the full λK-calculus are more subtle and did not come until much later.

The finiteness-of-developments theorem for λK was first proved by Schroer in his unpublished thesis [Schroer, 1965, Part I, Thm. 6.20]. It was re-proved by others using various methods, for example [Hyland, 1975, Thm. 3.6], [Hindley, 1978a, Thm. 1], and [Vrijer, 1985]. A proof by Barendregt, Bergstra, Klop and Volken in [Barendregt *et al.*, 1976, pp.14–17] used a particularly neat labelling technique and a strong normalization theorem for labelled reductions, see [Barendregt, 1981, §11.2].

Labelling techniques to follow components of terms down through reductions were first used in [Hyland, 1976, §2] and [Wadsworth, 1978, §4], in proofs of the limit theorem for approximants, see §9.2 below. In [Wadsworth, 1976, p.508] the idea was attributed by Wadsworth to a 1972 communication from Martin Hyland. It was extended by Jean-Jacques Lévy in 1974 [Lévy, 1976, pp. 106–114] and in [Lévy, 1978, Ch. II], and applied to prove the completeness of "inside-out" reductions and the other main theorems on reductions. A good general account of labelling is in [Barendregt, 1981, Ch. 14].

The standardization theorem for λK first appeared in [Curry and Feys, 1958, §4E1], though with a proof that depended on the finiteness of leftmost-first developments. In 1975 Curry's proof was much simplified by Gerd Mitschke, [Mitschke, 1979]. A consequence of standardization is the cofinality of leftmost reductions, i.e. that the leftmost reduction of a term is finite if and only if the term has a normal form. In the early 1970s Barendregt pointed out that the reductions most often used in practice were quasi-leftmost, not leftmost, and proved the corresponding cofinality theorem for these reductions too [Barendregt, 1981, Thm. 13.2.6].[43] In his 1976 report with Bergstra, Klop and Volken the more general concept of *reduction strategy* was introduced and studied, see [Barendregt *et al.*, 1976, pp.33–43] or [Barendregt, 1981, Ch. 13].

The basic equivalence theorem for complete developments in λK was first stated and proved in [Curry and Feys, 1958, pp.113–130]. In 1978 it was improved by Lévy to a stronger equivalence, see [Lévy, 1978, pp. 37, 65] or [Barendregt, 1981, §§12.2.1, 12.2.5]. Lévy's equivalence theorem led to a new view of β-reductions, embodied in the thesis [Klop, 1980]. The transformations involved in proofs of confluence and standardization were seen as conversions of *reduction diagrams*, and the reasonably tidy underlying structure that had been sought for the theory of β-reduction for the previous forty years was at last achieved. (An account is in [Barendregt, 1981, Chs. 11–14].)

For $\lambda\beta\eta$, confluence and the postponement of η-steps were known since Curry's

[43] A reduction is quasi-leftmost iff, for every non-leftmost step, some later step is leftmost.

work in the 1950s, as was mentioned in §5.3, but no standardization theorem had been proved analogous to that for β, nor had any cofinality theorem been proved for leftmost reductions, despite attempts. The first successful proofs of these theorems were made in [Klop, 1980, Ch. IV].

Returning to $\lambda\beta$: problems involved in making reduction more efficient began to attract attention in the 1970s, stimulated by programming needs. One such problem was that of finding a reduction of a term to its normal form that was optimal in some reasonable sense. For CL this problem was studied in 1972 by John Staples, though his work was not published until 9 years later, [Staples, 1981]. In 1974 for a system of recursive programs, Jean Vuillemin formulated a "delay rule" for reducing first-order terms with function symbols defined by a recursive program, [Vuillemin, 1974a, §3]. This rule in Vuillemin's system could be proved optimal and easily implementable, but the problem of optimality turned out to be more difficult in the case of λ-calculus. However, Lévy extended Vuillemin's idea to λ in his thesis [Lévy, 1978], by an in-depth analysis of the lattice of β-reductions starting at a term, and by exploiting his labelling technique to define certain families of redexes that can be shared through reductions.[44] Lévy also proved the completeness of *inside-out* β-reductions (almost the opposite of leftmost reductions), which had been conjectured in [Welch, 1975], see [Lévy, 1976].

Substitution was another topic which aroused interest from a computing point of view. Its correct definition had been settled years ago, but its efficient implementation now became an important problem; for example in LISP in the 1950s some of the details had been over-simplified and had had to be corrected, see §6.1 above. With the implementation problem in mind, in 1978 Gyorgy Révész proposed an interesting new alternative to β-reduction. It generated the usual equivalence relation $=_\beta$, was confluent, had the same normal forms as the usual β-reduction, and broke substitutions into smaller steps; see [Révész, 1978] or [Révész, 1985].

Another approach to the problem of implementing substitution efficiently was to extend the λ-language by introducing an *explicit substitution* operator with suitable reduction rules. This idea, which had occurred independently to Mitschke and Barry Rosen for use in confluence proofs for $\lambda\beta$-reduction ([Mitschke, 1973, pp. 150–151], [Rosen, 1973, §9]), led in the 1990s to several formal systems which are still being developed; a few examples of papers on these are [Komori, 1990], [Abadi *et al.*, 1991], [Kamareddine and Nederpelt, 1993], [Curien *et al.*, 1996], [Benaissa *et al.*, 1996], [Kamareddine and Rios, 1997], [Bloo and Geuvers, 1999], and [Kamareddine and Rios, 2000].

A special case of substitution was the problem of changing bound variables efficiently. In 1972 de Bruijn described a formalism which avoided this problem completely, by numbering all λs in a term and replacing bound variables by number-indices to show which λ bound them, [Bruijn, 1972]. De Bruijn's formal-

[44] A suitable data structure for implementing the sharing required for optimal reductions was described in 1990, independently in [Lamping, 1990] and [Kathail, 1990]. Then in [Gonthier *et al.*, 1992] and [Asperti, 1995] the technique was related to Girard's *geometry of interaction* [Girard, 1989a].

ism was used in the compiler of the programming language ML in the 1980s, and has been used in many explicit substitution systems.

8 TYPES

8.1 The general development of type theories

In the type theory of Russell and Whitehead's *Principia Mathematica* and its simplifications by Chwistek and Ramsey [Russell and Whitehead, 1913; Chwistek, 1922; Ramsey, 1926], there was little agreement as to what kind of entities types should be. Already in [Russell, 1903, §497] we find at least two ways in which types can be intended:

(1) *Types as ranges of significance of propositional functions:* given a propositional function $\varphi(x)$, there is a class of objects, the *type* of x, such that φ has a value whenever it is applied to a member of this type.

(2) *Types as sets:* individuals form the lowest *type* of objects, then there are sets of individuals, sets of sets of individuals and so on, ad infinitum.

The first interpretation of types is related to the theory of *grammatical categories*. The view of types as *sets* underlies, of course, the early developments in set theory, cf. [Quine, 1963, Ch. XI], but most notably from a λ-viewpoint it underlies the definition of the hierarchy of simple types in [Church, 1940] and much of the research on set-theoretical models of type theories.

The later grafting of the type-theoretic tradition onto λ and CL, and the mathematical study of the models of the resulting theories, contributed to the discovery of new ways in which types could be interpreted. A new unifying paradigm that became (and still is) very influential in research in type theory was introduced by Lawvere in 1969 and Lambek in the 1970s [Lawvere, 1969a; Lawvere, 1969b; Lambek, 1974; Lambek, 1980b; Lambek, 1980a]:

(3) *Types as objects in a category*, especially a *cartesian closed category*, whose definition provides a general framework for a formulation of the set-theoretical bijection between functions of several arguments and unary functions with functions as values.

In addition, research in proof-theory and the semantics of intuitionistic mathematics showed that it was in many cases fruitful to view

(4) *Types as propositions:* in either λ or CL, terms of a given type play the role of codes for proofs of (the proposition coded by) that type.

We shall now follow the historical development of each of these threads in turn.[45]

[45] We shall not try to survey all of type-theory, but only the parts that are most relevant to λ and CL. In particular, we shall not discuss the type-concept implicit in ZF set theory.

8.1.1 Types as grammatical categories

The theory of categorial grammars began in about 1900 with the attempt of Edmund Husserl to develop a theory of pure, *a priori* laws that underlie the composition of meanings, [Husserl, 1922, Fourth Investigation]. Through the teaching of Kazimierz Twardowski, Husserl's ideas influenced the Lwów school of logicians, and in particular the design of the logical systems of Leśniewski [Leśniewski, 1929], where they replaced the logical types of Russell. Ajdukiewicz was the first to give a self-contained formalization of a system of categorical grammar, [Ajdukiewicz, 1935]. There, it is assumed that to each word of a language can be assigned one or more *indices*, that are symbolic representations of the grammatical categories to which the word belongs. Starting from indices n for noun phrases and s for propositions, one can obtain infinitely many new indices α/β for phrases that yield a phrase of index α whenever prefixed to a phrase of index β. The concatenation of two words with respective indices α and β has index $\alpha\beta$, and one is allowed to replace each subsequence of the form $(\alpha/\beta)\beta$ by α. This basic system can be extended by forming the index $\alpha\backslash\beta$ for phrases that yield a phrase of index β whenever suffixed to a phrase of index α, with a rule that allows a subsequence of the form $\alpha(\alpha\backslash\beta)$ to be replaced by β, [Bar-Hillel, 1953; Bar-Hillel, 1959; Lambek, 1958; Lambek, 1961]. Writing $x \to y$ to mean that every expression of category x has also category y, one can prove theorems like, for example, that $xy \to z$ implies $y \to x\backslash z$. Lambek introduced a Gentzen-like calculus for deriving theorems of this form, and proved its decidability through a cut-elimination theorem, [Lambek, 1958].

Curry, because of his interest in the formal study of natural languages, knew about the theory of categorial grammar,[46] and noticed that his theory of functionality (see §4.2) had a grammatical interpretation in which types could be regarded as grammatical categories. This was in harmony with his early ideas on the theory of functionality, which was intended to be a tool for drawing categorial distinctions between the terms of systems of illative combinatory logic, with the idea that "a theory of types is essentially a device for saying that certain combinations are not propositions", see [Curry, 1980, §9]. The grammatical interpretation of functionality was described in [Curry, 1961];[47] it regarded the combination $\mathbf{F}\alpha\beta$ as the category of those functors that take an argument of category α and give a result of category β, in the context of a formal system whose only operation is application. (For example, if nouns and noun-phrases have category N, then adjectives have category $\mathbf{F}NN$, adverbs (as modifiers of adjectives) have category $\mathbf{F}(\mathbf{F}NN)(\mathbf{F}NN)$, and the suffix "-ly" (which changes adjectives to adverbs) has category $\mathbf{F}(\mathbf{F}NN)(\mathbf{F}(\mathbf{F}NN)(\mathbf{F}NN))$.)

The possibility of describing grammatical categories by means of types has also played a role in formalizing the syntax of logical theories within typed λ-calculi. This has happened, for example, with the theory of expressions that underlies Martin-Löf's intuitionistic theory of types (described for the first time by Martin-

[46] In particular, about [Ajdukiewicz, 1935], see the remark in [Curry and Feys, 1958, §8S1].

[47] Based on [Curry, 1948], summarized briefly in [Curry and Feys, 1958, §8S2, pp.274–275].

Löf in his lecture at the Brouwer Centenary in 1981; see [Nordström *et al.*, 1990, p.13]) and the more recent Edinburgh Logical Framework [Harper *et al.*, 1993]. In a broader perspective, the same connection has been exploited in mathematical linguistics in the area known as *Montague grammar*, [Montague, 1974].

8.1.2 Types as sets

There was a clear set-theoretic intuition behind the hierarchy of simple types in [Church, 1940]: Church used a type-expression $\beta\alpha$ to denote a set of functions from the set (denoted by) α to the set (denoted by) β. This intuition can readily be extended to yield a notion of model for the simply typed λ-calculus as a set-theoretic structure $\mathfrak{D}$ consisting of a family of sets $\{D^\alpha\}_{\alpha\in\mathbb{T}}$, where $\mathbb{T}$ is the set of all type-expressions of the system, such that $D^{\beta\alpha}$ is the set of all functions from D^α to D^β, with a family of application-functions $\cdot_{\alpha,\beta} : (D^{\beta\alpha} \times D^\alpha) \to D^\beta$, for $\alpha, \beta \in \mathbb{T}$, that correspond to the application of a function of type $\beta\alpha$ to an argument of type α with result of type β. This notion is essentially that of *standard model* introduced by Leon Henkin in [Henkin, 1950, pp.83–85], who also considered *general models* where $D^{\beta\alpha}$ is only assumed to be *some* class of functions from D^α to D^β.[48]

The interpretation of types as sets led to developments that became important also for later extensions of this rather limited framework. One of these was the notion of *logical relation* between models $\mathfrak{D}$ and $\mathfrak{E}$, defined as a family of relations $R^\alpha \subseteq D^\alpha \times E^\alpha$ indexed over types, such that[49]

$$R^{\alpha\to\beta}(f,g) \iff \big((\forall x \in D^\alpha, y \in E^\alpha)\ R^\alpha(x,y) \Rightarrow R^\beta(f \cdot_{\alpha,\beta} x,\ g \cdot_{\alpha,\beta} y)\big).$$

An early instance of the use of logical relations was the proof of completeness of typed $\beta\eta$-conversion with respect to any standard model (over an infinite set of individuals) of simply typed λ-calculus, given by Harvey Friedman in 1970 and published in [Friedman, 1975]. This result exploited the properties of *partial homomorphisms* of models, where a partial homomorphism from $\mathfrak{D}$ to $\mathfrak{E}$ is defined as a family of partial surjections $f^\alpha : D^\alpha \rightharpoonup E^\alpha$ from $\mathfrak{D}$ to $\mathfrak{E}$ for each type α, with the property that $f^\beta(x \cdot_{\alpha,\beta} y) \simeq f^{\alpha\to\beta}(x) \cdot_{\alpha,\beta} f^\alpha(y)$. The graph of a partial homomorphism is then a logical relation from $\mathfrak{D}$ to $\mathfrak{E}$.

Later research showed that logical relations afford a convenient framework for formulating and proving the main properties of simply typed λ-calculi. On the syntactic side, the various forms of the notion of "computability" used in Tait-style normalization proofs can be described as logical relations, as well as some other properties of reduction like confluence and η-postponement, [Statman, 1985]. On the semantic side, the extensional collapse of an applicative structure uses logical partial equivalence relations (a result from 1971 of Jeff Zucker, reported in [Troelstra, 1973, §2.4.5]; see also [Statman, 1983]). Special classes of logical

[48] Church's system had an extensionality axiom; for systems without this, the definition of model would be more complicated than in [Henkin, 1950]; see, for example, [Barendregt, 1981, 2nd edn., Ch.5] or [Hindley and Seldin, 1986, §11A].

[49] From now on we write the nowadays-common notation $\alpha \to \beta$ instead of Church's $\beta\alpha$.

relations were used in the characterization of definability in D_∞ models of the type-free λ-calculus and in standard models of the simply-typed λ-calculus by [Plotkin, 1973] (where the name 'logical relation', attributed to Mike Gordon, seems to have first appeared). See also [Plotkin, 1980].

The research on higher-order notions of computability in the 1950s produced a host of type structures or even models for simply typed λ in which function types are interpreted by some set of *computable* function(al)s.[50] Especially important from our standpoint is the search for a suitable notion of *partial* computable functional, leading to the partial *monotone* functionals studied in Richard Platek's Stanford PhD thesis [Platek, 1966] (see [Longley, 2001, §4.1.1] and [Moldestad, 1977]).

The idea of an ordering on functionals, stressed by Platek, together with the mathematical theory of computation of McCarthy [McCarthy, 1963a], are at the basis of the theory of computable functions of higher type described in a privately circulated typescript written in October 1969 by Dana Scott [Scott, 1969e] (see also [Scott, 1969d]). Scott proposed there a typed alternative to the (type-free) programming languages introduced in [Böhm, 1966; Böhm and Gross, 1966] and [Landin, 1966b], which amounted to a simply typed λ-calculus with ground types ι for integers and o for truth values, and containing a functional $\mathbf{Y} : (\tau \to \tau) \to \tau$, for all types τ, for forming (least) fixed-points of functions of type $\tau \to \tau$ (suggested by Platek's thesis), and McCarthy's conditional operator. In addition, there were constants Ω_ι and Ω_o denoting the "undefined" integer and truth value, respectively, so that partial numerical functions could be described as continuous total functions that may take value Ω_ι at some argument. The intended model for this system was based on the interpretation of types as *domains*, i.e. partially ordered sets D with a least element $\bot_D$ and least upper bounds of increasing chains. Functions over domains were required to be *continuous*: these are the functions that preserve least upper bounds of chains. The judgements of the system exploited the semantical ordering between values: notably, there was a rule of induction of the form:

$$\frac{\Phi \ \vdash \ g(\Omega) \leq h(\Omega) \qquad \Phi, g(x) \leq h(x) \ \vdash \ g(f(x)) \leq h(f(x))}{\Phi \ \vdash \ g(\mathbf{Y}(f)) \leq h(\mathbf{Y}(f))}.$$

This rule was considered by Scott to be the main advantage of the system (Scott [Scott, 1969e, §3]), as compared e.g. to the principle of recursion induction arising from the work of McCarthy [McCarthy, 1963a, §8].

The resulting system was the basis of the LCF project (LCF stands for "Logic for Computable Functions", the name given to Scott's system by Robin Milner) developed from 1972 onward by a group at Stanford University led by Milner. The functional programming language ML in [Gordon *et al.*, 1979] was originally developed in order to check proofs in a version of LCF, see [Gordon, 2000]. Further-

[50] For a thorough historical account of this field, highlighting its many intersections with our present subject, we refer the reader to [Longley, 2001]. Longley (ibid., §3.3.1) also points out [Kreisel, 1959] as the source of many ideas that have played a key role in the later development of λ.

more, the typed λ-calculus that constitutes its logic-free part became the language PCF ("Programming language for Computable Functions", [Plotkin, 1977]), that is widely used as the prototype functional programming language for many theoretical purposes, in particular the study of the important property of full abstraction (see later, §10).

8.1.3 Types as objects

The connection between type theory and category theory that underlies a large amount of current research shows up along the lines set by the works of Joachim Lambek and F. William Lawvere.

The latter realized in the early 1960s that certain basic mathematical transformations of a general nature in what he later called *categorical dynamics* [Lawvere, 1979] were in fact also basic structural ingredients of several other categories of foundational interest. In fact, these transformations were those used later to define the notion of cartesian closed category [Eilenberg and Kelly, 1966].

Though Lambek did not stress the category theoretic content of his results on the syntactic calculus in [Lambek, 1958; Lambek, 1961], we mention his ideas here because they represent an alternative route to the applications of cartesian closed categories to λ-calculus via deductive systems and a related notion of functional completeness, inspired by the abstraction algorithms of combinatory logic (§3 above). Furthermore, he later explicitly described formally the connection of the syntactic calculus with closed categories [Lambek, 1968; Lambek, 1988; Lambek, 1995], providing, with hindsight, the unifying framework for his work and that of Lawvere.

Lambek used the system of categorial grammar described in §8.1.1 as a formal tool in the construction of canonical mappings between functors of bimodules over associative rings.[51] There, the following correspondence between the constructors of syntactic types and operation on bimodules was exploited:

$$\begin{aligned} AB &=_{\text{def}} A \otimes_S B \\ C/B &=_{\text{def}} \text{Hom}_T(B, C) \\ A\backslash C &=_{\text{def}} \text{Hom}_R(A, C) \end{aligned}$$

for an R-S-bimodule A, an S-T-bimodule B and an R-T-bimodule C. There is an evident closed category structure in this example, in the form of what was called a *residuated category* in [Lambek, 1968]. In fact, already in his first published paper on the syntactic calculus, [Lambek, 1958, §10], Lambek noted the formal similarity between the operation of (right and left) division in the calculus of

[51] This application seems to have been one main mathematical motivation for the development of the syntactic calculus. It originated from work carried out by Lambek jointly with G. D. Findlay in 1955, described in the manuscript [Lambek and Findlay, 1955] (unpublished, but summarized in [Lambek, 1961, Appendix II]). According to [Lambek, 1988, p.297], the relation of this formalism to natural language was already clear to him at an early stage, but it was only later that, through a bibliographic search, he discovered the paper [Bar-Hillel, 1953] on categorial grammars.

syntactic types and residuation in lattice theory, the operation normally used to interpret implication in algebraic logic.

The notion of cartesian closed category emerged gradually, basically through the work of Lawvere who in 1959 became aware of the central role of exponential objects B^A, characterized by the isomorphisms $C \times A \to B \cong C \to B^A$.[52] An early example of this transformation came from the description of the motion of a material body seen as a mapping $T \times M \to E$ that gives the position of each of its particles at each time instant, or equivalently, as its trajectory $T \to E^M$ through time, or as the path $M \to E^T$ followed by each of its particles, [Lawvere, 1976, p.123].

There are other instances of exponentiation that played an important role in the development of the notion of cartesian closed category. The introductory chapter of Lawvere's PhD thesis [Lawvere, 1963] contains the description of the category of categories, where the exponential of two objects (namely, categories) $\mathcal{A}$ and $\mathcal{B}$ is the category of all functors from $\mathcal{A}$ to $\mathcal{B}$.[53] Later, in the years 1963–64, Lawvere investigated an elementary[54] axiom system for the category of sets in order to develop a basic framework for analysis, [Lawvere, 1964]: the exponential B^A of two objects A and B is here the set of all functions with domain A and codomain B. The relation of exponentiation to λ-abstraction is mentioned for the first time in this work, but no formal detail is given. During his stay in Berkeley in 1961–62, Lawvere had noted that Heyting algebras are cartesian closed categories: more generally, a preordered set with finite products has a monoidal structure which is closed if and only if the preorder is residuated, that is, iff it also satisfies the equivalence:

$$p \wedge q \leq r \text{ if and only if } p \leq q \Rightarrow r,$$

where $\Rightarrow$ is the residuation operation. (This example appears in [Eilenberg and Kelly, 1966, Chapter IV, §4]). This is one of the possible ways of arriving at what was later called the "propositions as types" correspondence (§8.1.4).

A general notion of category $\mathbb{C}$ with exponentiation, that took the name of *cartesian closed category* after discussions with Eilenberg and Kelly [Eilenberg and Kelly, 1966, Chapter IV, §2], was described by Lawvere by the following adjoint situations [Lawvere, 1969a]:

- a terminal object, given by a functor $1 : \mathbf{1} \to \mathbb{C}$ right adjoint to the unique functor from $\mathbb{C}$ to the category $\mathbf{1}$ with one object;
- a bifunctor $\times : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ (cartesian product) right adjoint to the diagonal functor $\Delta : \mathbb{C} \to \mathbb{C} \times \mathbb{C}$ that transforms an object C of $\mathbb{C}$ into the object $\langle C, C \rangle$ of $\mathbb{C} \times \mathbb{C}$; and finally

[52] Intuitively, these may be thought of as the tranformation of a two-argument mapping $f : C \times A \to B$ into the one-argument mapping $\hat{f} : C \to B^A$ which outputs a function as value. See §3, currying.

[53] This was later published as [Lawvere, 1966].

[54] In this context, "elementary" refers to the fact that the axioms are sentences of a multi-sorted first-order language—for example, with one sort for maps and another for objects.

– a right adjoint $(\cdot)^A : \mathbb{C} \to \mathbb{C}$ to the functor $A \times (\cdot) : \mathbb{C} \to \mathbb{C}$, for each object A of $\mathbb{C}$. The counit of each adjunction $A \times (\cdot) \dashv (\cdot)^A$ is a family of morphisms indexed by objects B of $\mathbb{C}$, $\varepsilon_B : B^A \times A \to B$, called *evaluation*.

The explicit connection between cartesian closed categories and type theory was highlighted in Lawvere's unpublished manuscript "Category-valued higher-order logic", distributed in 1967 to the Los Angeles Set Theory Symposium, whose contents were later published as [Lawvere, 1969a; Lawvere, 1969b; Lawvere, 1970a]. The central categorical notion in these papers is that of a *hyperdoctrine*, and was intended to capture some of the examples that would later lead to the notion of *elementary topos* [Lawvere, 1970b]. A hyperdoctrine is defined by a cartesian closed category $\mathbb{T}$ of *types*, whose morphisms are regarded as *terms*. For each type X in $\mathbb{T}$, there is a cartesian closed category $P(X)$ of *attributes of type* X, whose morphisms are called *deductions over* X. The terminal object of $P(X)$ is regarded as the identically true attribute of type X, whereas cartesian product and exponentiation represent conjunction and implication of attributes, respectively, evaluation being *modus ponens*. For every term $f : X \to Y$ in $\mathbb{T}$, there are adjoint situations

$$\sum_f \quad \dashv \quad f \cdot (\quad) \quad \dashv \quad \prod_f$$

that correspond to existential quantification, substitution along f, and universal quantification, respectively.

Examples of hyperdoctrines include, of course, the syntactical hyperdoctrine whose types are the type symbols of a formal type theory and whose attributes of type X are the formulas whose free variables have types corresponding to X. More interesting from the present point of view (see also later, §8.1.4), is an example suggested by the work of H. Läuchli [Läuchli, 1970], in which a complete semantics for intuitionistic predicate calculus is described in terms of a set-theoretic notion of construction inspired by Kleene's notion of realizability and the theories of constructions of Goodman, Kreisel, Troelstra and Scott.[55] In this example $\mathbb{T}$ is the category **Set** of sets, and for any set X seen as a type, the category of attributes of type X is $\mathbf{Set}/X$, so that each attribute φ of type X may be regarded as an X-indexed family of sets, whose x-th coordinate is written as $x \cdot \varphi$. A noteworthy feature of this hyperdoctrine is the interpretation of the logical connectives in the category of attributes of type X, for example [Lawvere, 1970a]:

– $\varphi \supset \psi$ is a family of sets whose x-th coordinate is $(x \cdot \psi)^{(x \cdot \varphi)}$. Unfolding the definitions, this means that a proof over X of $\varphi \supset \psi$ is a function which, for every $x \in X$, assigns a proof of $x \cdot \psi$ to a proof of $x \cdot \varphi$;

[55] [Läuchli, 1970] appeared in the proceedings of a conference held at Buffalo in 1968, but an earlier manuscript was circulated already in the Spring of 1967 and was discussed by Lawvere and Dana Scott in San Francisco in the same year. See also [Lawvere, 1996], from which it appears that his manuscript "Category-valued higher-order logic" was discussed also at the symposium on Automatic Demonstration, where the two fundamental papers on the interpretation of proofs as typed λ-terms, [Scott, 1970a] and [Bruijn, 1970], were presented. We shall discuss later, in §8.1.4, the relation between the works of Lawvere and Scott.

– $y \cdot \sum_f(\varphi) = \sum_{f(x)=y} x \cdot \varphi$. This means that a proof over Y of $\sum_f(\varphi)$ is a function assigning to every $y \in Y$ an ordered pair $\langle x, d\rangle$ where $f(x) = y$ and d is a proof of $x \cdot \varphi$.

In the meantime, in a series of papers on "deductive systems and categories", Lambek was exploiting the description of categories by means of equational theories over what he called *deductive systems* [Lambek, 1968; Lambek, 1969; Lambek, 1972]. These are essentially (oriented multi)graphs with an edge $1_A : A \to A$ for every node A and an edge $g \circ f : A \to C$ for every pair of edges $f : A \to B$ and $g : B \to C$. In the case of cartesian closed categories, the right kind of deductive system has the following additional structure [Lambek, 1974; Lambek, 1980a]:

– a node T with an edge $0_A : A \to T$ for each node A;

– a node $A \times B$ for every pair of nodes A, B, with edges $\pi_{A,B} : A \times B \to A$ and $\pi'_{A,B} : A \times B \to B$ and $\langle f, g\rangle : C \to A \times B$ for every pair of edges $f : C \to A$ and $g : C \to B$;

– a node B^A for every pair of nodes A, B, and edges $\varepsilon_{B,A} : B^A \times A \to B$ and $\hat{h} : C \to B^A$ for every edge $h : C \times A \to B$.

Observe that a deductive system as described here is neutral as to the interpretation of its nodes and edges as formulas and proofs of the positive intuitionistic propositional calculus or as objects and morphisms of a cartesian closed category, respectively. Therefore, equations between edges of a deductive system can be interpreted either as describing a notion of equivalence of proofs for intuitionistic sequents[56] or as the equational presentation of a cartesian closed category.

The equational presentation of cartesian closed categories allowed Lambek to give a smooth proof of their *functional completeness* [Lambek, 1972; Lambek, 1974], thus making fully explicit the connection with typed λ-calculus already hinted at by Lawvere. If $\mathcal{D}$ is a deductive system, then an assumption of a formula A may be regarded as an *indeterminate*, i.e. as a new edge $x : T \to A$. The edges $\varphi(x) : B \to C$ of a deductive system $\mathcal{D}[x]$ obtained by adding to $\mathcal{D}$ the indeterminate x can be interpreted either as proofs of an intuitionistic sequent $B \to C$ depending on the assumption x, or as polynomials in a cartesian closed category. The following property, proved in [Lambek, 1974, Th. 3.6], subsumes therefore both the deduction theorem of the positive intuitionistic propositional calculus and the functional completeness of a simply typed λ-calculus with finite products (and surjective pairing):

[56] The theme of equivalence of proofs is a recurring one in Lambek's works: it was already a motivation for his study of the syntactic calculus, leading to a technique for proving coherence theorems in category theory [Kelly and Mac Lane, 1972] based on [Lambek, 1968], and also opened the way for the categorical analysis of proofs pursued in [Szabo, 1974; Szabo, 1978] and [Mann, 1975].

With every $\varphi(x) : B \to C$ in $\mathcal{D}[x : T \to A]$ there is associated in $\mathcal{D}$ an $f : A \times B \to C$ such that

$$f \circ \langle x \circ 0_B, 1_B \rangle \ =_x \ \varphi(x) : B \to C,$$

where $=_x$ is the extension of the equivalence of edges of $\mathcal{D}$ to $\mathcal{D}[x]$.

Functional abstraction $\lambda x : A \,.\, \varphi(x)$ is then just defined to be $\widehat{\psi} : B \to C^A$, where $\psi = f \circ \langle \pi'_{B,A}, \pi_{B,A} \rangle$. One of the outcomes of the categorical formalization of type theories consisted in providing a way of speaking of types without necessarily assuming that they are sets in a strict sense, because categories with a rich enough structure have an internal logic that still allows one to reproduce set-theoretical arguments to a significant extent. This was exploited systematically by Lawvere in his work on elementary topoi, regarded as universes of "variable sets", [Lawvere, 1970b; Lawvere, 1972; Lawvere, 1976].

This insight set the stage for the most advanced developments in domain theory, especially the idea of synthetic domain theory (§10 below) as a comprehensive framework unifying constructive mathematics and computation theory.

8.1.4 Types as propositions

The interpretation of types as objects of a cartesian closed category, and the observation that such a category corresponds precisely to positive intuitionistic propositional calculus, is only one way of arriving at the analogy between types and formulas that is nowadays called the "propositions as types" correspondence. As mentioned in §4.2, the first published hints of this correspondence date back to [Curry, 1934a] and [Curry, 1934b], and it was first described explicitly in [Curry and Feys, 1958, §9E], as an isomorphism between the provable formulae of intuitionistic implicational logic and the type-expressions assignable to closed terms.

(In the early 1950s, Carew Meredith had observed that proofs in this logic corresponded to composite combinators, and that certain fragments of this logic corresponded to restricted sets of terms, but his observations were not published till later, see §7 of *Notes on the axiomatics of the propositional calculus* by C. A. Meredith and A. N. Prior, *Notre Dame Journal of Formal Logic* 4 (1963), 171–187.)

In 1969 Curry's correspondence was extended by Howard to one between *proofs* in a sequent calculus version of intuitionistic logic and typed λ-terms. He further observed (after [Tait, 1965]) that cut-elimination was a consequence of the normalization of terms. Howard's 1969 manuscript circulated informally and became the *locus classicus* of the correspondence, especially in the proof-theoretic literature, after [Prawitz, 1971] and [Girard, 1972]. It was eventually published in [Howard, 1980]. (Roughly, the idea behind Howard's contribution may be described as a generalization of Gödel's theory of primitive recursive functionals of finite type by extending the set of types to formulas of Heyting arithmetic.)

Independently of these proof-theorists and with very different motivations, but at about the same time, there were also two other efforts that led to formalisms

centred systematically around the correspondence between proofs and terms, and whose technical and philosophical insights still underlie current work on applications of typed λ-calculi to the mechanization of mathematical proofs. These were Dana Scott's work on "constructive validity" [Scott, 1970a], motivated explicitly by an attempt to analyze the notion of *construction* as used by the intuitionists, and Nicolaas de Bruijn's Automath family of languages for checking the correctness of ordinary mathematical proofs, [Bruijn, 1980].

The starting point of the Automath project was around 1967, when de Bruijn began to look at the problem of devising a formal language such that the correctness of a completely formalized mathematical proof can be verified by a computer program [Bruijn, 1970].[57] The design of the Automath languages did not depend on any philosophical position concerning the foundations of mathematics; in fact, it is possible to describe the overall format of an Automath text (called a *book* in the Automath jargon) by asking what items of information a machine would need in order to check its correctness. First of all, the text is subdivided into *lines*, each of them composed of the following items: (1) an *identifier* to refer to that line later in the book; (2) a *definition*, indicating whether the identifier is a *primitive notion*, a *free variable* or the abbreviation of a compound *expression* formed from previous identifiers; (3) a *context* that keeps track of which assumptions are valid and which identifiers are currently visible at that line, and finally (4) a *category* for the identifier, including a primitive category type.[58] The context structure of the Automath languages is a way of describing proofs in a natural deduction format, and was suggested by the similarity between the block structure of the programming language Algol [Naur and others, 1963] and the way mathematicians structure proofs.[59] The λ-notation was familar to de Bruijn, who used it systematically in his work in analysis, and it was natural for him to include (typed) λ-terms in Automath in order to discharge assumptions by binding free variables, thus making contexts shrink as well as expand.

More relevant from the point of view of this Section is the interpretation of the category part of an Automath line. The leading formal insight here consists in unifying the treatment of proofs and objects by observing that in informal mathematical discourse the two notions play the same role. In the statement of a theorem T it is usual to declare the category of the objects involved, e.g. "Let x be an integer," and the properties that these objects have to satisfy, e.g. "Assume that $x > 0$." The conclusion of the theorem then states a property that depends

[57] Our description here is based on [Bruijn, 1994], which contains several remarks of historical interest on the Automath languages. More details can be found in [Bruijn, 1980], and the editors' introduction to [Nederpelt *et al.*, 1994], which is a collection of the basic papers on Automath by de Bruijn and others.

[58] For a heuristic explanation of how the format of Automath lines can be motivated starting from a complete formalization of natural deduction proofs, see [Daalen, 1980]. Observe that, except for the identifier part, the components of a line correspond to those of an ordinary typing judgement $\Gamma \vdash e : C$, for a context Γ, an expression e and a category C.

[59] See [Bruijn, 1994]; the natural deduction formalism used by de Bruijn is close to that used by [Fitch, 1952]. It is worth remarking that the block structure of Algol 60 led Landin to the formal analysis of that language by means of λ-calculus [Landin, 1965].

on the (free) name x, which can be instantiated to any specific object k provided that k is an integer *and* that we have a proof that $k > 0$. Then the theorem T is more faithfully represented by an expression of the form $\mathrm{T}(k, \pi)$, showing its dependence on both k and a proof π that $k > 0$. In a line where the identifier a has the expression e of category C as *definiens* (in a context Γ), a has two possible interpretations: as the name of the object e of type C, or as the name of a proof e of a proposition P, provided that C can be interpreted as the type $\mathsf{Proof}(P)$ of proofs of P.[60] Types are not formally identical to propositions; rather, to every proposition can be assigned a type, the type of its proofs, and asserting a proposition amounts to saying that the type of its proofs is inhabited [Bruijn, 1970].[61] This paradigm of "proofs as objects" specializes to the constructive interpretation of logical constants following [Heyting, 1956] where, for example, a construction of an implication $A \supset B$ is a function that maps each construction of A to a construction of B. This is recognized in [Bruijn, 1994, p.205] to have been one of the clues to the interpretation of proof classes as types.[62]

Typed λ-abstraction is denoted, in languages of the Automath family, by $[x\!:\!A]B$ where B might be a type. If, in a context Γ extended with the assumption that $x\!:\!A$, the expression e (possibly depending on x) has type B, then the abstraction $[x\!:\!A]e$ has type $[x\!:\!A]B$; but observe that here the expression B can also depend on the variable $x\!:\!A$ (i.e., B may contain free occurrences of x). One interpretation of the type $[x\!:\!A]B$ is set-theoretic: it is the type of functions that assign a result of type $B(a)$ to every argument a of type A, a type that may be written as $\prod_{x:A} B(x)$.[63] When B does not depend on x, $[x\!:\!A]B$ is just another notation for the function type $A \to B$. Another interpretation is logical: if A and B represent the proof classes $\mathsf{Proof}(P)$ and $\mathsf{Proof}(Q)$, respectively, then the type $[x\!:\!\mathsf{Proof}(P)]\,\mathsf{Proof}(Q)$ represents the proof class $\mathsf{Proof}(\forall x\!:\!P.\,Q)$, where the formula $\forall x\!:\!P.\,Q$ specializes to $P \supset Q$ when Q does not contain free occurrences of x.

A more uniform approach is obtained in later languages of the Automath family, [Daalen, 1973], where besides the category type there is a category prop intended to represent the type of proofs of propositions. If $A\!:\!\mathsf{prop}$ and $B\!:\!\mathsf{prop}$ whenever $x\!:\!A$, then $[x\!:\!A]\,B$ can be read as an *abbreviation* of $\forall x\!:\!A.\,B$ and, when $x\!:\!A$ is not free in B, of $A \supset B$.

A different and more complex background was behind [Scott, 1970a]. Scott was interested in setting up a formal calculus of constructions that would substantiate Heyting's claim that "there is no essential difference between logical and mathematical theorems, because both sorts of theorems affirm that we have succeeded

[60] In de Bruijn's papers, this type was denoted by $TRUE(P)$.

[61] This is similar to Kreisel's notion of *modified realizability* [Kreisel, 1959] for HA^{ω}, where the type of realizers of a formula $A \supset B$ is the function-set $\sigma \to \tau$, if σ is the type of realizers of A and τ of those of B. See [Oosten, 2002] for a history of realizability.

[62] The constructive explanation of the meaning of logical constants was learned by de Bruijn directly from Heyting, who was one of his colleagues in Amsterdam from 1952 to 1960.

[63] If B contains x free, this is a *dependent type*; for other examples, see §8.3 and **G** in §5.4.

in performing constructions satisfying certain conditions" [Heyting, 1958].[64]

The work [Scott, 1970a] lay at the confluence of several threads. On the one hand Kreisel's papers [Kreisel, 1962; Kreisel, 1965] aroused Scott's interest in the problem of finding a theory of proofs and constructions on which the interpretation of intuitionistic logic could be based. Kreisel's project was taken up by Goodman [Goodman, 1970] (based on his PhD thesis in Stanford in 1968, supervised by Dana Scott), who deviated from the original formulation in allowing partial functions, represented by type-free combinators.

On the other hand, there was the work of Läuchli [Läuchli, 1965; Läuchli, 1970] on a set-theoretical version of Kleene's recursive realizability interpretation of intuitionistic logic [Kleene, 1945], that we have seen in §8.1.3 above to have influenced also Lawvere's work in categorical logic.[65]

Scott took as central the connection, already exploited by Läuchli and Lawvere, between the universal quantifier and the cartesian product of an indexed family of species, and considered also the dual notion of the sum of such a family, with the evident connection to existential quantification. In general terms, Scott's system was "an attempt at axiomatizing in a constructive way a theory of *both* functions and families of sets of functions" [Scott, 1970a, p.241].[66] The system of Scott uses intuitionistic sequents $\Delta \vdash \mathcal{A}$ where formulas $\mathcal{A}$ are of the two shapes $A \in B$ and $A = B$, for terms A, B that may denote functions or species depending on the context, where species here should be understood as *types*. Term-forming operations include typed functional abstraction $\lambda x \in A.\, B$, which satisfies standard rules with the notable exception of rule (ξ) to allow for an intensional interpretation. The type of the above λ-abstract is a dependent product given by the rule:

$$\frac{\Delta,\ x \in A \ \vdash\ B \in C}{\Delta \ \vdash\ (\lambda x \in A\,.\,B) \ \in\ (\forall x \in A\,.\,C)} \quad (\text{where } x \notin \Delta)$$

which is at the same time the $\forall$-introduction rule. The dual elimination rule is

[64] A detailed description of the motivations behind [Scott, 1970a], on which our account is based, is contained in that paper on pp.237–242. The development of Scott's theory of constructions was also influenced at an early stage by the design of Automath: a preliminary version of the above paper, containing a discussion of de Bruijn's language, was presented in November 1968 at Troelstra's seminar [Scott, 1968], while Scott was on sabbatical leave in Amsterdam. The paper itself was read in December of the same year at the Versailles conference on Automatic Demonstration, whose proceedings also include the first published paper on Automath, [Bruijn, 1970].

[65] The Russian logician Ju. T. Medvedev in [Medvedev, 1962] had described an interpretation of the positive part of the intuitionistic propositional calculus in terms of finite problems, following Kolmogorov, that is quite close to that given by Läuchli. In particular, Medvedev interpreted conjunction, disjunction and implication by means of cartesian product, disjoint union and exponentiation of sets, as Läuchli did.

[66] The attempt to develop a purely category-theoretic treatment of parameterized families of sets started around 1970 in Lawvere's lectures at Dalhousie University, was further developed in [Lawvere, 1972] and led to the theory of indexed categories, [Johnstone *et al.*, 1978], that may be seen as a continuation of Lawvere's earlier work on hyperdoctrines (§8.1.3). A related approach was that of [Bénabou, 1985] using fibrations, and both approaches are relevant to the semantics of dependent types and higher-order type theories discussed later, in §8.3.

the typing rule for application:

$$\Delta,\ f \in (\forall x \in A\,.\,B),\ x \in A \ \vdash\ f(x) \in B.$$

The usual (non-dependent) function type becomes $A \to B =_{\text{def}} \forall x \in A\,.\,B$ where B does not depend on $x \in A$. It is under this identification that the correspondence between types and propositions becomes evident, and in fact a theorem of this calculus of constructions like

$$x \in A \ \vdash\ [\forall f \in (\forall x \in A\,.\,B)\,.\,f(x)] \ \in\ (\forall x \in A\,.\,B) \to B$$

is the logical axiom of universal instantiation [Scott, 1970a, p.248]. The clearest statement of how the correspondence works in general is to be found in this paper (p.260):

> "We began by regarding species and constructions as mathematical objects and found that there were some simple axioms governing their properties. It then became slowly apparent that these properties were highly analogous to properties familiar from formal logic. We then turned this analogy into a dogma by insisting that the logical formulas be read (better: interpreted) as (mathematically meaningful) terms of the theory of constructions. This interpretation requires that validity be asserted by the act of giving an explicit construction belonging to the (interpretation of the) formula. Validity is established by giving a proof from the axioms for constructions of the membership assertion. [...] a proposition does not simply degenerate to one of two truth values but instead is represented by a complex species of possible constructions that conceivably can be used in its validation."

Scott's full description of the correspondence follows almost *verbatim* the explanations of the meaning of the logical constants given in [Heyting, 1956; Heyting, 1958]:

Formula	*Construction*
$A \wedge B$	a pair of constructions, the first of which justifies A, and the second B
$A \vee B$	a pair whose first coordinate is either 0 or 1: if 0 then the second coordinate justifies A; if 1, then B
$\top$	the justification is 0, as $\top =_{\text{def}} \{0\}$
$\bot$	no justification is known, as $\bot =_{\text{def}} \{\ \ \}$
$\forall x \in A\,.\,B$	a construction that maps every element $a \in A$ into a justification of $B(a)$
$\exists x \in A\,.\,B$	a pair whose first coordinate is an element $a \in A$ and whose second coordinate justifies $B(a)$.

The correspondence between proposition and types was used systematically in the intuitionistic type theory of Martin-Löf. That theory was first proposed in 1971

but then underwent several revisions; cf. [Martin-Löf, 1971b; Martin-Löf, 1972b; Martin-Löf, 1975; Martin-Löf, 1984].[67] Martin-Löf had learned about this correspondence from Howard in early December 1968, and had exploited it already in a paper, written in March 1969 and eventually published as [Martin-Löf, 1972a], on the infinitary proof theory implicit in [Tait, 1965]. The first proposal of intuitionistic type theory included a type constant V such that $A : V$ meant that A is a type, with the axiom that $V : V$. This assumption allowed for a very succinct system: the primitive notions could be reduced, essentially, to typed λ-abstraction, dependent product and the constant V, using the coding of Russell and Prawitz to define the following basic type constructors (i.e., logical connectives) [Russell, 1903, §§18,19], [Prawitz, 1965, Ch. V, Thm. 1]:

$$\begin{aligned} \bot &=_{\text{def}} \forall X : V \,.\, X \\ A \times B &=_{\text{def}} \forall X : V \,.\, (A \to B \to X) \to X \\ A + B &=_{\text{def}} \forall X : V \,.\, (A \to X) \to (B \to X) \to X \\ \exists x : A \,.\, B(x) &=_{\text{def}} \forall X : V \,.\, (\forall x : A \,.\, (B(x) \to X) \to X. \end{aligned}$$

There were substantial motivations for making the strong impredicative assumption of a type of all types, based on the simultaneous acceptance of

> "the following three principles. First, quantification over propositions, as in impredicative second order logic. Second, Russell's doctrine of types according to which the ranges of significance of propositional functions form types so that, in particular, it is only meaningful to quantify over all objects of a certain type. Third, the identification of propositions and types" [Martin-Löf, 1971a, p.5].

However, the theory turned out to be inconsistent, as a form of the Burali-Forti paradox could be coded in it.[68] The impredicative character of the theory was removed in its subsequent versions [Martin-Löf, 1972b; Martin-Löf, 1975; Martin-Löf, 1982; Martin-Löf, 1984]. Broadly, the outcome can be described as an intuitionistic theory of iterated inductive definitions developed in the framework of dependent types,[69] but here the correspondence between types and propositions is even more systematically exploited than in the systems discussed earlier. The correspondence suggests a formulation of the rules for type-forming operations as introduction and elimination rules, as in [Gentzen, 1935]. This yields

[67] Martin-Löf's work on this theory began in October 1970. It grew out of earlier proof-theoretical work on normal form theorems (*Hauptsätze*) for the intuitionistic theory of iterated inductive definitions, the theory of species and intuitionistic simple type theory, and also an interest in category theory, which was found inadequate as a foundational framework because of its lack of a clear distinction between extensional and definitional equality.

[68] This was proved by Girard [Girard, 1972, III, Annexe A], and his paradox was studied in [Coquand, 1986]; for a study of this system from a different point of view, see [Jacobs, 1989], who suggested that Girard's paradox could be viewed as a version of the paradox of Mirimanoff on the universe of well-founded sets [Mirimanoff, 1917].

[69] This appropriate description was given in [Coquand and Dybjer, 1994].

automatically a notion of reduction for terms as deductions as in [Prawitz, 1965; Prawitz, 1971], from which computation rules can be read off directly. The resulting system then becomes a functional programming language including its own logic, whose formulas are regarded as program specifications, as Martin-Löf himself pointed out in [Martin-Löf, 1982].[70] It was that paper that first outlined a correspondence between proof-theory and the theory of functional programming languages that had a widespread conceptual and terminological influence on logically-oriented functional programming. In particular, under the identification of propositions with types, the introduction rules behave as the *constructors* of [Landin, 1964], while logical constants correspond to operations that form new types by specifying what are their canonical elements (i.e., those whose outermost symbol is a constructor), like the *concrete data types* of functional programming languages since [Burstall, 1977] (see [Peyton Jones, 1987, §4.1.4]).

On the more philosophical side, these ideas are also in accordance with a theory of meaning that started from [Gentzen, 1935, §II.5.13], according to which the meaning of a logical constant is determined by its introduction rules in a system of natural deduction, whereas the corresponding elimination rules depend on this meaning. It was this idea that led to a precise formulation in [Prawitz, 1965] of the "inversion principle" of [Lorenzen, 1955], where it contributed to isolating the reduction rules for derivations in natural deduction. The meanings of the constants of Martin-Löf's intuitionistic theory of types were first studied in [Hancock and Martin-Löf, 1975] for a simple language of primitive recursive definitions; these studies were brought to a more mature state with [Martin-Löf, 1984; Martin-Löf, 1985; Martin-Löf, 1987], and were very relevant to the work on the semantics of intuitionistic logical constants, and more generally to the constructive theories of meaning and the philosophical foundations of the intuitionistic interpretation of mathematical propositions (see [Sundholm, 1986] for a survey of this area until 1985).

Work related in several ways to Martin-Löf's theory of types had also been carried out at Cornell University by Robert Constable and his collaborators from the early 1970s. Constable attributed to Kleene's work on realizability in the 1940s the awareness that various kinds of constructive proofs can be compiled into executable code, [Constable, 1982, p.3]. For instance a proof of an assertion of the form "for all integers n there is an integer y such that $R(n, y)$" can be compiled into a computable function f such that $R(n, f(n))$.

He proposed in [Constable, 1971] that certain constructive formal systems could be used directly as programming languages, and designed a programming logic based on this idea in [Constable and O'Donnell, 1978]. Further refinements of this system, [Constable, 1980; Constable, 1985; Bates and Constable, 1985], together with the inclusion of ideas from [Martin-Löf, 1982] and from Edinburgh LCF (§8.1.2), led eventually to the Nuprl system [Constable and others, 1986] that

[70] The particular technique used in describing the computation rules in [Martin-Löf, 1982] is closely related to what is now called "structured operational semantics", cf. [Plotkin, 1981] and [Kahn, 1988].

manipulated proofs directly in a version of Martin-Löf's intuitionistic type theory.

From the researchers in the Nuprl group also emerged an extension of the "types as propositions" idea to classical logic. In particular, Tim Griffin and Chetan Murthy ([Griffin, 1990], [Murthy, 1991]) observed a correspondence between the *continuation passing style* translation of control operators for λ-calculus [Felleisen *et al.*, 1987] and the double-negation translation of classical into intuitionistic logic as described, for example, in [Friedman, 1978].

8.2 Early normalization proofs

From a syntactic point of view, the most desirable properties of a system of typed terms include the *weak normalization (WN)* property that every typed term has a normal form, and, better, the *strong normalization (SN)* property that no infinite reductions exist.

The first proof of such a property for λ or CL was by Turing, as mentioned in §5.1. Turing's proof covered λ-terms with pure arrow types and β-reduction. It was found in a typewritten note among his papers after his death, and is known to have been composed before 1942, but was not published until about 40 years later, when it appeared as [Turing, 1980].

In the meantime Curry made a very different WN proof in the 1950s using cut-elimination, and this was the first proof to be published; it appeared in [Curry and Feys, 1958, §9F6, Cor. 9F9.2].[71]

The method used by Turing was re-discovered independently several times in the 1960s and '70s, for example by James H. Morris in his unpublished thesis [Morris, 1968, p.107 Thm. 2], and by Garrel Pottinger in [Pottinger, 1978, p.448] (with ordinals added in). The first published proof using Turing's method was in [Andrews, 1971, p.417, Prop. 2.7.3], where the author's source for the proof was given as James Guard, who was, like Andrews, a 1960s student of Church, and the theorem was attributed to "folklore".

For stronger type-systems, several WN proofs were made in the 1960s. The initial stimulus was Gödel's *Dialectica* paper [Gödel, 1958], together with studies of the ideas behind it led by Kreisel in Stanford University. Roughly speaking, Gödel interpreted first-order arithmetic in a logic-free theory T of *primitive recursive functionals of finite type*, and deduced the consistency of arithmetic by finitary reasoning from an assumption that all closed terms in T of numerical type (closed type-0 terms) computed to unique numerals. One way of formalizing T was as a version of typed CL or λ augmented by primitive-recursion operators, and the above computability property of T then became the statement that WN held for all closed type-0 terms of the augmented CL or λ..[72] This drew the problem of

[71] Curry's WN theorem was for CL with $\beta\eta$-strong reduction. It was deduced via a λ-CL translation from his cut-elimination theorem for a system based on $\lambda\beta$, see §5.4 above.

Its proof seems to have been made independently of Curry's co-author Robert Feys; no sign of any input from Feys was found in a recent search of Curry's detailed notes by J. P. Seldin, [Seldin, 2008, §2.6].

[72] A simple formalization of Gödel's T is the system $(\mathrm{CL}{+}\mathbf{R})_t$ defined in outline in [Hindley and

verifying the WN theorem for such extensions of CL and λ to the forefront, and at least five different WN proofs came into being by 1966. Two were made around 1963 by Tait. One of these was published in [Tait, 1965]; it covered a version of λ augmented by certain infinite terms, but based on a weak λ-reduction analogous to weak reduction in CL. The other was reported in the informally circulated notes of a seminar held at Stanford in 1963 (Section V Appendix B), but was not actually published until [Tait, 1967]; it covered CL augmented by operators for primitive recursion and bar-recursion. It introduced the method due to Tait that is now more or less standard, of defining by a carefully organised induction on types a *computability predicate* that implies normalizability, and proving by induction on terms that all terms satisfy this predicate.[73]

Other early normalization proofs for versions of Gödel's T included ones in 1965 by Luis Sanchis for $(\mathrm{CL}{+}\mathbf{R})_t$ [Sanchis, 1967, Thm. 8], in 1966 by Justus Diller for $(\lambda\beta{+}\mathbf{R}')_t$ [Diller, 1968, §6], in 1966 by Yoshito Hanatani [Hanatani, 1966], in 1967 by Shigeru Hinata [Hinata, 1967], and in 1968 by A. G. Dragalin [Dragalin, 1968]. For more on normalization proofs stemming from Gödel's T, see [Troelstra, 1973, §§2.2.1–2.3.13, esp. 2.2.35].

The property SN can be seen as a safety feature for computing systems. But at first there was no interest in this property, as WN was all that was needed in consistency proofs; the WN proof in [Tait, 1967] would have needed almost no change to give SN had Tait wanted it. The first explicit SN proof was Sanchis' 1965 proof mentioned above. An SN proof by directly assigning ordinals to terms was made by Howard in 1968, [Howard, 1970]; it worked for weak λ-reduction augmented by $\mathbf{R}$, but not for full-strength $\lambda\beta\mathbf{R}$-reduction. (An extension to that reduction had to wait for nearly 30 years, until Gunnar Wilken and Andreas Weiermann's [Wilken and Weiermann, 1997].) After the 1960s, most workers studying normalization aimed to prove SN whenever possible.

Normalization proofs for Gentzen-style natural deduction systems of logic are also relevant here, because deductions in such systems can be viewed as a kind of typed λ-terms. Some such proofs will be mentioned in the following sections. For others, the reader is referred to the history of mainstream proof theory.

Seldin, 1986, 1st edition, pp.297–299]. It consists of CL with weak reduction plus a constant $\mathbf{R}_\tau$ for every type-expression τ, with reduction-rules $\mathbf{R}_\tau XY\mathbf{0} \triangleright X$, $\mathbf{R}_\tau XY(\mathbf{succ}\,\mathbf{n}) \triangleright Y\mathbf{n}(\mathbf{R}_\tau XY\mathbf{n})$. (Types are omitted here, and $\mathbf{0}$ and $\mathbf{succ}$ are atoms.) A proof of WN (indeed SN) for $(\mathrm{CL}{+}\mathbf{R})_t$ is in [Hindley and Seldin, 1986, 1st edition, Appendix 2], based on the method in [Tait, 1967].

By the way, although $\mathbf{R}_\tau$ was adequate to formalize T, [Tait, 1967] and many other formalizations were actually equivalent to a slightly stronger operator $\mathbf{R}'_\tau$ with the rule $\mathbf{R}'_\tau XY(\mathbf{succ}\,Z) \triangleright YZ(\mathbf{R}'_\tau XYZ)$.

[73] In [Tait, 2001, §5] Tait expressed the opinion that his computability idea had been anticipated by Gödel in unpublished notes in 1938 and lectures which were given in 1941 but not published in print, see [Gödel, 1995, items *1938a, *1941]. But our interest here is only in explicit proofs of WN, and none occurs in these sources.

8.3 Higher-order type theories

In 1953, Gaisi Takeuti formulated a sequent calculus GLC for higher-order logic and conjectured that the cut-elimination property would hold for it, i.e. that if a formula had a derivation D in GLC then it would have a derivation D^* in which the cut-rule was not used, [Takeuti, 1953].

After 13 years, two proofs of Takeuti's conjecture were eventually found for the second-order fragment of GLC [Tait, 1966; Prawitz, 1967], and Moto-o Takahashi and Dag Prawitz found proofs for the whole of GLC [Takahashi, 1967; Prawitz, 1968]. But these proofs gave no procedure for finding D^* from D; they all relied on a result of Kurt Schütte which reduced the conjecture to the problem of proving that every partial valuation could be extended to a total valuation [Schütte, 1960].

The first proof to contain a procedure to find D^* was made by Jean-Yves Girard in 1970, see [Girard, 1971; Girard, 1972].

Girard formulated two systems of typed λ-calculus, $\mathbb{F}$ corresponding to the second-order fragment of GLC, and the system nowadays called $\mathbb{F}_\omega$ corresponding to full GLC, and he solved Takeuti's problem via proofs of SN for these systems. The key idea in his SN proofs, *candidats de réductibilité*, was a development of Tait's computability method, [Tait, 1967], and was very quickly adopted and extended by other workers in proof-theory and has been used widely since.[74]

Girard's system $\mathbb{F}$ is particularly important in our story.[75] Its main novelties were the possibility of forming universally quantified types $\forall t.\sigma$, and the operation of applying a term M of type $\forall t.\sigma$ to a type τ. Corresponding to this operation there was an abstraction operation $\Lambda t.M$ to represent a function whose application to a type τ reduced to the term $M\{\tau\}$ of type $[\tau/t]\sigma$. The impredicativity implicit in this extension, whereby the type variable t in $\forall t.\sigma$ ranged over *all* types, including $\forall t.\sigma$ itself, made the SN proof for reductions in this calculus far from obvious.

The impact of system $\mathbb{F}$ was limited to proof theory in the years immediately following its invention, but the idea of having λ-terms taking types as parameters was also becoming important in those years in the theory of programming languages (see, for example, [Cheatham Jr. *et al.*, 1968]). It is not surprising, then, that almost the same system was independently reinvented in 1974 by John Reynolds, [Reynolds, 1974].[76] One motivation for the development of a system of quantified types was the formalization of ideas on polymorphism in programming languages expressed by Christopher Strachey (see, e.g., [Strachey, 1967, §3.6.4]), especially the notion of *parametric polymorphism*. Take, for example, the function map that applies a function f to all elements of a list. If $f : \sigma \to \tau$, then it is natural to assign to $\mathsf{map}(f)$ the type $\mathsf{List}(\sigma) \to \mathsf{List}(\tau)$ that depends parametrically

[74] An analysis of Girard's method, and of several SN proofs that use it, is [Gallier, 1990].

[75] For an introduction to system $\mathbb{F}$, see [Girard *et al.*, 1989].

[76] Reynolds has often pointed out that he presented his paper in April 1974 at a meeting at Paris VII University, where Girard was teaching, but nobody in the audience noticed the similarity of the two systems or brought Girard's work to his attention.

on σ and τ. By using universally quantified types, map could be assigned the type

$$\forall s.\forall t.\ (s \to t) \to (\mathsf{List}(s) \to \mathsf{List}(t)).$$

An important remark was that polymorphic terms can be used to represent basic data structures; for example the natural numbers can be represented by polymorphic Church numerals of the form

$$\Lambda t.\ \lambda f : t \to t.\ \lambda x : t.\ \ f(\cdots(fx)\cdots),$$

each having type

$$\mathsf{nat} \ = \ \forall t.\ (t \to t) \to (t \to t),$$

see [Girard, 1972, p. I.5.5]. Representations which occurred in [Girard, 1972, pp. I.5.2, I.5.3], also in [Martin-Löf, 1971b, §3] which referred to [Russell, 1903], were

$$\begin{aligned} \sigma \times \tau &= \Lambda t.\ ((\sigma \to (\tau \to t)) \to t), \\ \sigma + \tau &= \Lambda t.\ ((\sigma \to t) \to ((\tau \to t) \to t)). \end{aligned}$$

The boolean values had representations

$$\Lambda t.\ \lambda x : t.\ \lambda y : t.\ x, \qquad \Lambda t.\ \lambda x : t.\ \lambda y : t.\ y$$

of type

$$\mathsf{bool} \ = \ \forall t.\ t \to (t \to t).$$

The representation of a list $[e_1, \ldots, e_n]$ of elements of some type s appeared in [Reynolds, 1983, p.520] as the term

$$\Lambda t.\ \ \lambda f : s \to (t \to t).\ \ \lambda x : t.\ \ fe_1(\cdots(fe_n x)\cdots)$$

of type

$$\mathrm{list}(s) \ = \ \forall t.\ (s \to (t \to t)) \to (t \to t)$$

and was said to have been "suggested by a functional encoding of lists devised by C. Böhm".

Another source for representations by polymorphically typed terms was [Leivant, 1983]; this had some influence on Böhm and Berarducci [Böhm and Berarducci, 1985] which contained a systematic account of how to encode many-sorted term algebras as closed types of the polymorphic λ-calculus, together with the iteratively defined functions over them as polymorphic terms.[77]

We have remarked that the main motivation for Reynolds' development of the polymorphic typed λ-calculus was the formal clarification of Strachey's notion of *parametricity*, by describing the uniform dependence of a polymorphic term $\Lambda t.M$ upon an input type. Informally, the behaviour of a parametric function defined by

[77] There are several accounts of these representations in the literature: for example [Girard *et al.*, 1989, §§11.3,11.4,11.5] and [Leivant, 1990, §3], besides the original papers.

such a term should not depend on the input type or, in other words, should be the same for all input types [Reynolds, 1983, p.519]. The main technical difficulty was to compare values whose types were different, and Reynolds proposed a semantic criterion of parametricity, consisting in the property of terms of the form $\Lambda t.M$ of preserving certain hereditarily defined relations between sets corresponding to different types [Reynolds, 1983].[78] Reynolds' notion of parametricity was based, however, on a set-theoretical interpretation of polymorphic types extending the general models of Henkin (§8.1.2 above). Accordingly, the relational account of parametricity showed several similarities with the theory of logical relations, in particular with the treatment of definability in [Plotkin, 1980].

The existence of a set-theoretic model for the polymorphic typed λ-calculus was conjectured in [Reynolds, 1983]. But a year later Reynolds disproved his own conjecture by showing that the functor (over the category of sets and functions) $(2^X)^X$ is definable by a type expression, hence should have an initial algebra $A \cong (2^A)^A$, which is impossible by a cardinality argument.[79]

The model theory of system $\mathbb{F}$ was initiated by Girard in his [Girard, 1972, §II.5], which contained the description of a model HEO_2 (and its generalizations to higher-orders HEO_n) based on the Kleene applicative structure $\langle \mathbb{N}, \cdot \rangle$, where $n \cdot m$ denoted the result of applying the n-th partial recursive function to the argument m. Every type σ was interpreted as a pair consisting of a set $D_\sigma \subseteq \mathbb{N} \times \{\sigma\}$ and an equivalence relation $=_\sigma$ over D_σ.

For functional types $\sigma \to \tau$ this interpretation was specified by the clauses:

- $D_{\sigma\to\tau}$ is the set of all pairs $(e,\ \sigma \to \tau)$ such that for all (e_1, σ) and (e_2, σ) in D_σ, $e \cdot e_i$ is defined for $i = 1, 2$ and is an element of D_τ, and $(e \cdot e_1, \tau) =_\tau (e \cdot e_2, \tau)$ whenever $(e_1, \sigma) =_\sigma (e_2, \sigma)$;

- $(e_1, \sigma \to \tau) =_{\sigma\to\tau} (e_2, \sigma \to \tau)$ iff $(e_1 \cdot x, \tau) =_\tau (e_2 \cdot x, \tau)$ for all (x, σ) in D_σ.

The interpretation of for-all-types used the notion of *primary type* P, defined as a set $A \subseteq \mathbb{N}$ containing 0 together with an equivalence relation on A. Then the pair $D_{\forall t.\sigma}$ with the equivalence $=_{\forall t.\sigma}$ was specified by the clauses:

- $D_{\forall t.\sigma}$ is the set of all pairs $(e, \forall t.\sigma)$ such that $(e, [P/t]\sigma) \in D_{[P/t]\sigma}$ for all primary types P;

- $(e_1, \forall t.\sigma) =_{\forall t.\sigma} (e_2, \forall t.\sigma)$ iff $(e_1, [P/t]\sigma) =_{[P/t]\sigma} (e_2, [P/t]\sigma)$ for all primary types P.

[78] Formal systems based on this relational approach to parametricity are described in [Abadi *et al.*, 1993] and [Plotkin and Abadi, 1993]. For a survey of parametricity see also [Longo *et al.*, 1993, §1] and [Longo, 1995].

[79] Reynolds' result was generalized by Plotkin in a letter to Reynolds dated July 16, 1984. His remarks led to [Reynolds and Plotkin, 1990]. (By the way, [Reynolds, 1974] contained a proposal for a model of Reynolds' first polymorphic system, but with a warning that its discussion contained a "serious lacuna". As he has pointed out since, this warning turned out to be justified, in that the proposed model did not work.)

With some simplifications, see [Troelstra, 1973, §§2.4.11, 2.9.7], this was to become probably the best-known model for the polymorphic typed λ-calculus, the model based on partial equivalence relations (*PERs*), i.e., symmetric and transitive relations over some type-free structure like $\langle \mathbb{N}, \cdot \rangle$ or, later, over some model of untyped λ-calculus [Scott, 1976, §7].[80]

Reynolds' proof of the non-existence of set-theoretic models of polymorphism made an essential appeal to classical logic. In fact, two years after Reynolds' result, it turned out that, within a constructive metatheory provided by the internal logic of an elementary topos, the set-theoretic intuitions behind the polymorphic λ-calculus could be rescued by internalizing the constructions of category theory within such a topos, looking at the latter as a constructive universe of sets [Pitts, 1987].

Already in 1984, Eugenio Moggi described to Martin Hyland, then visiting Pisa, a way of looking at a model of polymorphic λ-calculus as an internal cartesian closed category $\mathbb{M}$ in a topos E, with $\mathbb{M}$ having all products indexed over the objects of $\mathbb{M}$ itself. Moggi found such a model in 1986 by taking E to be the effective topos of [Hyland, 1982] and $\mathbb{M}$ as the category of *modest sets*, which was equivalent to the familiar category of partial equivalence relations over the applicative structure $\langle \mathbb{N}, \cdot \rangle$, directly related to Girard's HEO_2 model of system $\mathbb{F}$. The right notion of completeness needed to make $\mathbb{M}$ a small complete category was then found by [Hyland, 1988].

Remarkably, the intersection of a family of partial equivalence relations becomes their product under internalization. It is the existence of such a small complete category inside a universe of constructive sets that enables one to interpret a polymorphic type $\forall X.A[X]$, roughly, as a product $\prod_{X \in \mathbb{M}} A_X$ of a family of elements of $\mathbb{M}$ indexed by $\mathbb{M}$ itself.[81] In particular, completeness of $\mathbb{M}$ implies that every functor $T : \mathbb{M} \to \mathbb{M}$ has an initial algebra $\langle A,\, a : TA \to A \rangle$, which is the stumbling block to the construction of a set-theoretic model of system $\mathbb{F}$.

From the present point of view, it is of some interest to observe that these developments exploit in an essential way the language of category theory, especially the notion of *indexed category* [Johnstone *et al.*, 1978], which has become a fundamental tool in the semantical description of theories of dependent types and may be regarded as a natural outgrowth of Lawvere's work on hyperdoctrines and topoi in the 1970s.

A polymorphic type $\forall t.\sigma$ is, implicitly, a dependent type $(\Pi t\!:\!\mathsf{type}).\,\sigma$, provided suitable rules about the constant type are added to systems with dependent types like those of §8.1.4 or §5.4. As mentioned in §5.4, a general system of dependent types was developed and studied from 1972 to 1975 by Seldin, [Seldin, 1979]. But a stronger system offering an integration of polymorphism with dependent types

[80] Some historical remarks on the notion of PER as an interpretation of types are given in [Bruce *et al.*, 1990], where we learn that they were introduced in [Myhill and Shepherdson, 1955] for types of first-order functions, and then extended to simple types by [Kreisel, 1959].

[81] This result was announced in a famous message by Moggi to the `types` electronic forum, at that time moderated by Albert Meyer, dated February 10, 1986. The result and its background are discussed in much more detail in [Longo and Moggi, 1992] and [Asperti and Longo, 1991].

was the *calculus of constructions*, developed by Thierry Coquand and Gérard Huet from 1984 onwards. It was described in a series of papers starting with Coquand's Thèse de Troisième Cycle: [Coquand, 1985], [Coquand and Huet, 1985], [Coquand and Huet, 1988], [Coquand, 1990]. The calculus of constructions has been concisely described as "a version of type theory that expresses naturally Heyting semantics of intuitionistic higher-order logic" [Coquand, 1990, p.91]. Paradoxes like that described by Girard for Martin-Löf's impredicative version of intuitionistic type theory were avoided by adopting a restricted form of the identification between types and propositions, which only assumed that to each proposition is associated the type of its proofs. But the other two points in Martin-Löf's argument for introducing a type of all types quoted on page 764 above, namely quantification over propositions and the doctrine of types as ranges of significance of propositional functions, were preserved. The resulting system was therefore quite similar to that described in [Martin-Löf, 1971b], and also to some of the systems in the Automath family, extended by quantification over elements of type, a possibility already envisaged in [Bruijn, 1970, §12.6]. For the calculus of constructions a weak normalization proof was given in [Coquand, 1985, §4, Cor. 1], together with some hints on how Girard's notion of candidat de réductibilité might be extended to give an SN proof for this system. A proof of SN for deductions and for terms was included in [Seldin, 1987, §4.3] and [Seldin, 1997, §3].

8.4 Intersection types and recursive types

In systems where types denote sets, a natural extension of the language of simple type theory would be to adjoin types of the form $\sigma \cap \tau$ to denote the intersection of whatever sets were denoted by σ and τ.[82] This extension was actually made in the late 1970s. Its history was, however, less simple than might be expected given the naturalness of the idea.

Systems of intersection types are usually "Curry-style" systems, in which types are assigned to untyped terms by rules and an infinite number of types may be assigned to one term, rather than "Church-style" systems, in which each term has a unique built-in type. They also use types, not as safety-devices to avoid paradoxes, as Russell and Church did, but simply as labels to describe the formal computational behaviour of terms.

According to ideas of Böhm on the use of λ-calculus as a programming language (cf. §6.3), data and programs are represented by λ-terms thought of as states of a computation that evolves by applications of the rule of β-reduction, with normal forms corresponding to final states. The notion of program equivalence that arises from this approach regards two programs (i.e. λ-terms) P and P' as equivalent if, for every argument F, either PF and $P'F$ both have normal form and $PF =_\beta P'F$, or both fail to have a normal form.[83] In order to match this notion of equivalence with convertibility, it turned out to be essential that P, P'

[82] In the present subsection and §8.5, "type" is short for "type-expression".

[83] This is close to the notion of contextual equivalence in [Morris, 1968] discussed in §7.1.

and F have normal forms, see [Böhm and Dezani, 1975]. This remark prompted a search for a classification of normal forms according to their normalizability when applied to other normal forms.

Starting in 1975, a suitable classification was defined and analysed by Böhm and his former students in Turin, Mariangiola Dezani and Mario Coppo; see [Böhm and Dezani, 1975], [Dezani, 1975], and [Böhm *et al.*, 1977]. The theme of types was introduced in [Dezani, 1975]; each normal form was given a finite sequence $(m, j, h_1, \ldots, h_n)$ or $(\omega, h_1, \ldots, h_n)$, called its "type", which encoded a description of its computational behaviour. Types appeared again in [Coppo and Dezani, 1978], written in 1977, in which simple types were assigned to all normal forms, including those untypable in previous simple type theories, by new rules involving two formally defined relations between types, an equivalence and a pre-order. The atomic types were 0, to represent the set of all normal forms, and 1 for the set of all terms whose application to any finite sequence of normal forms always reduces to a normal form.

In 1977 Coppo and Dezani put forward the idea of assigning a finite sequence of types to a term to say that this term denoted a member of a finite intersection. This was published in [Coppo and Dezani, 1978], and the results in that paper included the theorem that the types received by a term were invariant under conversion in the λI-calculus.

The sequence idea was extended by Patrick Sallé in Toulouse, [Sallé, 1978, §5]. He adjoined a new atomic type-constant ω to denote the universal set, and by this means he extended the conversion-invariance theorem from λI to the full λK-calculus, and described the connection between non-ω types and head-normal forms, see [Sallé, 1978, Thms. 9–12] and the joint paper [Coppo *et al.*, 1979, Thms. 1, 2]. The system with sequences and ω was further studied in [Coppo *et al.*, 1980] and [Coppo *et al.*, 1981]. In §5 of the former paper, Betti Venneri showed that, by interpreting a closed term as the set of all its types, one could obtain a new model of untyped λ-calculus.

The intersection idea was also invented independently by Pottinger, who proved some of its basic properties in [Pottinger, 1980].

But although the intersection system was perfectly adequate for its original purpose, it was not semantically complete. Suitable extra deduction-rules to make it complete were first formulated in 1980, by Coppo, Dezani and Barendregt together, [Barendregt *et al.*, 1983], and by Hindley independently, [Hindley, 1982]. The system given in [Barendregt *et al.*, 1983] soon became standard: its types were built from variables and ω by means of constructors $\rightarrow$ and $\cap$, and it contained rules defining a relation $\leq$ between types analogous to the subset relation. In what is nowadays often called the "generation lemma" [Barendregt *et al.*, 1983, Lemma 2.8] the invertibility of its main rules was proved, and from this a normalization theorem for deductions was deduced, [Barendregt *et al.*, 1983, Cor. 4.9]. In [Barendregt *et al.*, 1983, §3] the first of the interesting class of λ-models called *filter models* was defined, whose elements were filters (i. e., $\leq$-upward closed and $\cap$-closed sets) in the set of all types. These models turned out to be closely related

to the original Scott construction to be discussed below in §9.1.

Over the next 10 years the system of [Barendregt *et al.*, 1983] was studied intensively, with many interesting results emerging, for example an analysis of its principal types by Simonetta Ronchi della Rocca and Venneri in [Ronchi della Rocca and Venneri, 1984] and [Ronchi della Rocca, 1988]. Applications of intersection types included their use in type systems for the control of interference in Algol-like languages, [Reynolds, 1989] and in static analysis of functional programs, [Coppo and Ferrari, 1993].

The use of types for describing behavioural properties of terms motivated several other type systems in the 1980s, for statically detecting properties of λ-terms and (functional) programs. Among these, the *recursive types* (see, e.g., [MacQueen *et al.*, 1984], [Coppo, 1985] and [Cardone and Coppo, 1991]) were a natural extension of Curry's type inference system by adjoining types of the form $\mu t.\sigma$ to represent the solution of the equation $t = \sigma$, where the type variable t may occur in σ. (The possibility of having circular type expressions was already considered in [Morris, 1968], see above §7.1.) Clearly, recursively typed λ-terms lack the normalization property, e.g. $\Delta\Delta$, where $\Delta \equiv \lambda x.xx$, can be assigned every type σ, provided x is given type $\mu t.t \to \sigma$ (which is equivalent to $(\mu t.t \to \sigma) \to \sigma$). But there is a principal typing algorithm for the basic type inference system for recursive types (extending fairly directly the one for Curry's system, §8.5), and this makes it a rather practical tool for preventing incorrect applications when special constants like e.g. arithmetical primitives are added to the λ-calculus [Wand, 1987; Cardone and Coppo, 1991].

The semantics of recursive types systematically exploited the techniques invented towards the end of the 1960s for solving recursive equations involving sets. Typical among these is the solution of the equation $D = D \to D$ satisfied by a model of pure, extensional λ-calculus that will be discussed at some length in §9.1.

Recursive types are also interesting when added to higher order type systems like system F, or to systems that include a subtyping relation: such a system was first proposed by Luca Cardelli and Peter Wegner in [Cardelli and Wegner, 1985] as a tool for analyzing certain features of object-oriented programming languages, originating a whole new field of research; see, for example, [Gunter and Mitchell, 1994] and [Pierce, 2002].

8.5 Algorithms for simple types

For a given system of terms and types, a *typability test* decides whether an untyped term M is typable in the system; it is called a *principal type algorithm* if it furthermore assigns to any typable term its principal type, i.e. the most general type it can receive under the system's rules. Such algorithms have no meaning for a Church-style type-system, which has no untyped terms in its language, but only for Curry's style, in which types are assigned to untyped terms by rules, cf. §4.2.

In §4.2 we mentioned that Quine's concept of *stratified formula* was based on a Curry-style approach, [Quine, 1937, pp. 78–79]. In 1942 a test for stratification

was invented by M. H. Newman and stratification was proved equivalent to typability, [Newman, 1943, §§1–4]. But Newman also extended his algorithm to give a typability test for λ-calculus, and included a proof of its correctness. Unfortunately his exposition was rather abstract, and his paper seems not to have been read by any later worker on typability tests until the preparation of the present history, as far as we know, see [Hindley, 2008].

Newman's typability test was not a principal type algorithm; indeed, his paper contained no concept of principal type. A more direct test, which moreover gave principal types, was described informally in [Curry and Feys, 1958, §§9B2–4]. Roughly speaking, for a given term M its method is to list all M's subterm-occurrences with their types given as type-variables, write out equations connecting these variables, and then solve these equations as generally as possible. Several examples were worked out in detail in [Curry and Feys, 1958], but the method was not stated explicitly as a formal algorithm there. Curry first wrote it out as a formal algorithm, with a proof of its correctness, in private notes [Curry, 1966] which led to a paper [Curry, 1969a].

A different principal type algorithm, which used the unification algorithm of [Robinson, 1965] to save itself some work, was made and proved correct by Hindley for CL in 1967, [Hindley, 1969b, §3]. (Curry and Hindley communicated during the writing of [Curry, 1969a] and [Hindley, 1969b] but deliberately kept their approaches different.)

A Curry-style equation-solving algorithm for λ was made, with a correctness proof, by James H. Morris in his thesis [Morris, 1968, Ch. 4, §E], independently of Curry's and Hindley's 1966–69 work.

Then the 1970s and '80s saw the development of the programming language ML by the group led by Robin Milner in Edinburgh. ML included a type-assignment system for λ plus the extra operator *let*, and Milner described a principal type algorithm for it in [Milner, 1978, §4.1, Algorithm W]. This algorithm was unification-based like that in [Hindley, 1969b], though it was invented independently. It was rewritten and extended in [Damas and Milner, 1982], and its correctness was proved by Luis Damas in his thesis [Damas, 1984].

After 1980, typability and principal-type algorithms were made for various more complex systems, some depending on pre-existing unification algorithms and some being self-sufficient: for a survey see [Tiuryn, 1990]. A long-standing open problem was the existence of a typability test for Girard's system $\mathbb{F}$ (see §8.3 above); this was raised by Daniel Leivant in [Leivant, 1983], and was solved in the negative by Joe Wells in 1994; indeed Wells proved that not even type-checking for this system was decidable, [Wells, 1994; Wells, 1999].

But the history of principal type algorithms is not confined to type theory, strange to say. Under the propositions-as-types correspondence, simple types whose atoms are variables become propositional formulas, and a combinator M with principal type τ becomes a proof of τ by the propositional rule called *condensed detachment*.

Roughly speaking, the condensed detachment rule corresponds to the construc-

tion of the principal type of a term PQ from those of P and Q. It was invented by the Dublin logician Carew Meredith at some time before 1954, and was first used in print in [Lemmon *et al.*, 1957, §9]. Meredith never published a formal statement of it, but his cousin David Meredith was taught the rule by him in 1954, wrote it up as a program, and ran it on the computer UNIVAC1 in the U.S.A. in 1957.[84] Thus, by 1957 the key step in a principal-type algorithm had not only been stated formally, but had been implemented on a machine.

Looking further back: an argument very like an analysis of a particular case of condensed detachment was used in [Łukasiewicz, 1939, Engl. edn., p.276]. Łukasiewicz attributed his method to Tarski, and in [Łukasiewicz and Tarski, 1930, Engl. edn., p.44] a method was mentioned which might perhaps be the one to which he referred. But the method was not actually described there, and "might perhaps" is conjecture not history. (See [Kalman, 1983, §4] for further comments.)

Furthermore, there is a big difference between an algorithm's key step and the whole algorithm with its necessary correctness-proof added. So, although there were partial precursors, and logicians as strong as Łukasiewicz and Tarski could surely have done more if they had wished to, we must conclude that the first typability algorithm was Newman's and the first principal-type algorithm was Curry's.[85]

Other algorithms for the simple type theory of pure λ and CL include the *counting algorithm* made by Choukri-Bey Ben-Yelles in 1979 for counting and listing the normal inhabitants of a given type, see [Ben-Yelles, 1979, Ch. 3] or [Hindley, 1997, Ch. 8].[86] Ben-Yelles' method was later modified by Sabine Broda and Luis Damas to give an algorithm which counts and lists just the principal normal inhabitants, [Broda and Damas, 1999].

Several algorithms for constructing a principal inhabitant of a type from any other inhabitant have also been published. The first was in [Hindley, 1969b, §6], for full λK. Others, which worked also for restricted sets of terms such as λI, were made by Robert Meyer in 1986, Grigori Mints and Tanel Tammet in 1988, and Norman Megill and Martin Bunder in 1994; see [Meyer and Bunder, 1988], [Meyer *et al.*, 1991, §4], [Mints and Tammet, 1991], [Megill and Bunder, 1996].

As David Meredith noted in [Meredith, 1980], such an algorithm provides a proof that the condensed detachment rule is complete for the weak implicational logic corresponding to the restricted λ-calculus in question.

---

[84] For a formal statement of the condensed detachment rule, see [Kalman, 1983, §2] or [Hindley, 1997, §7D]. For further historical information, see those references and [Meredith, 1977]. The rule has similarities with the well known resolution rule of J. A. Robinson, and pre-dates it.

[85] There is no sign that Curry's 1958 co-author Feys contributed to the invention of his algorithm, see Footnote 71 or [Seldin, 2008, §2.6].

[86] An *inhabitant* of a type τ is any closed term M with type τ. It is *normal* iff it contains no β-redexes. It is *principal* iff τ is its principal type.

9 MODELS FOR λ

9.1 Scott's D_∞ model

The first set-theoretical model for untyped λ was made by Dana Scott while he was in Oxford to work with Strachey during a leave from the Princeton Philosophy Department.[87] In [Scott, 1969c, p.51] he gives a precise date for the idea: 14 November 1969. Only one month before, he had strongly advocated a typed alternative to the (untyped) programming languages defined in [Böhm, 1966; Böhm and Gross, 1966] and [Landin, 1966b], in a privately circulated manuscript written for Strachey ([Scott, 1969e], mentioned in §8.1.2 above). His model was unexpected, and showed that despite his previous criticisms of untyped languages they could be given a mainstream mathematical sense. Further, the means he used for this turned out to be a natural outgrowth of the notions on which his earlier typed alternative had been based.

In the first versions of Scott's model construction, domains were complete lattices, [Scott, 1969b]. In a later account of the construction in [Scott, 1970c], he introduced *continuous lattices*, characterized algebraically as those complete lattices D where, for every $y \in D$,

$$y \;=\; \bigvee\{\bigwedge U \mid y \in U \text{ and } U \text{ is Scott-open and } U \subseteq D\},$$

or topologically as those T_0-spaces such that every continuous $f : X \to D$ from a subspace $X \subseteq Y$ can be extended to a continuous $\overline{f} : Y \to D$.[88] It was this extension property that made continuous lattices especially attractive for "a coherent theory of partial functions (in extension)" [Scott, 1973, p.178], and more precisely for "an extensional theory of functions, related directly to computability, where functions (rather than syntactic representations or programs for the functions) can be treated as objects in themselves" (Scott, in [Stoy, 1977, p.xxviii]).[89]

From a global point of view, domains and Scott-continuous functions formed cartesian closed categories where the exponential objects $[D \to E]$ were the complete lattices of Scott-continuous functions from D to E: this made them suitable for the semantics of typed λ-calculus (see §8.1.3). But they also contained *reflexive* domains D_∞, that solved the equation $D_\infty = [D_\infty \to D_\infty]$ non-trivially

[87] The history of this collaboration, and the circumstances in which the model was invented, are given in Scott's foreword to [Stoy, 1977], in his Turing Award lecture [Scott, 1977], in [Scott, 2000], and in his comments in the 1993 published version of [Scott, 1969e].

[88] See [Scott, 1972] for details. Here $\bigvee$ and $\bigwedge$ are the usual lattice operators. A Scott-open subset U of a domain D is an upward closed set such that if $\bigvee \Delta \in U$ for a directed $\Delta \subseteq D$ then $U \cap \Delta \neq \emptyset$. A function $f : D \to D'$ is (Scott-) continuous iff it is monotonic and preserves the least upper bounds of directed subsets of D. An instance of this topology had already been considered in [Nerode, 1957; Nerode, 1959], and even before, in [Uspenskii, 1955]. The interplay of algebraic and topological notions has been a fruitful one in recent developments of domain theory, see §10 later.

[89] The theory of continuous lattices gained further impetus from their rediscovery and application in analysis, functional analysis and topological algebra, as documented in [Gierz *et al.*, 1980]. Their connection with *interval analysis* was pointed out in [Scott, 1970b, p.87].

(up to isomorphism), and these yielded models of the untyped λ-K-calculus with $\beta\eta$-conversion.

The insight that led Scott to the construction of reflexive domains consisted basically in extending to operations on domains, in particular $\mathcal{F}(D) = [D \to D]$, the iterative construction of the least fixed-point of a continuous function $f : D \to D$ as

$$\mathrm{fix}(f) \quad =_{\mathrm{def}} \quad \bigvee_{n\in\omega} f^n(\bot). \tag{1}$$

The technical device needed for this extension was later recognized by Scott himself [Scott, 1973, p.181] to be the most original step of his construction: given domains D and E, an *embedding-projection pair* from D to E is a pair of (Scott-) continuous functions $i : D \to E$, $\ j : E \to D$ such that $j \circ i = \mathrm{id}_D$ and $i \circ j \sqsubseteq \mathrm{id}_E$, where the order relation on functions is defined pointwise. Embedding-projection pairs defined an approximation relation on domains in analogy to the approximation relation on elements of domains described by the partial ordering. Reflexive domains were built as inverse limits of sequences of projections

$$\langle D_n \overset{j_n}{\leftarrow} D_{n+1}\rangle_{n\in\omega}$$

where D_0 was arbitrary and D_{n+1} was defined as $[D_n \to D_n]$, with a suitable choice of the initial embedding-projection pair $\langle i_0, j_0\rangle$. These spaces were "function-space analogues to the simple types of the Russell-Zermelo theory" [Scott, 1969b, p.24], and Scott always regarded his models as spaces of functions of infinite type, rather than of untyped functions, [Scott, 1975, pp.348–349].[90]

Soon after Scott's D_∞ models, several other models were invented by other workers, see 9.3 below, and there was intensive research on the structure of D_∞ and its application to the denotational semantics of programming languages. Probably the first result in this area was due to David Park [Park, 1976], who proved in early 1970 that the interpretation of the Rosenbloom-Curry combinator $\mathbf{Y}$ in D_∞ coincided with the least fixed point operator defined iteratively as in (1) above. On the negative side, Park also showed that this definability result could be ruined by a different choice of the initial embedding-projection pair $\langle i_0, j_0\rangle$ from D_0 to D_1.

9.2 Computational and denotational properties

A subject of extensive research in the early 1970s was the relation between the operational properties of λ-terms and their denotations as elements of some D_∞ model. A particularly interesting problem was to relate the lattice structure of the models, where the order relation among elements reflected their amount of information in an abstract sense (see, for example, [Scott, 1970c, p.171]), to the

[90] Actually, in his first note on the construction, in [Scott, 1969a], Scott used direct limits of sequences of embeddings, obtaining an equivalent result. This was the first instance of what is now called the *limit-colimit coincidence* [Smyth and Plotkin, 1982], that lies at the heart of the categorical versions of the construction.

intuition that some terms are more informative than others in an operational sense. For example, there are λ-terms X such that, for all contexts $C[\ ]$ and all normal forms N,

$$\text{if } C[X] =_\beta N \text{ then } C[M] =_\beta N \text{ for every other term } M; \tag{2}$$

one of these is $\Delta\Delta$, where $\Delta \equiv \lambda x.xx$. Such terms contribute no information to the surrounding context $C[\ ]$, and could therefore be interpreted as the least element $\perp$ in any model. This problem was investigated thoroughly by Christopher Wadsworth, who was at the time a graduate student of Strachey at Oxford University, in his 1971 PhD thesis and two later papers, [Wadsworth, 1971; Wadsworth, 1976; Wadsworth, 1978].[91] He found a precise and easily grasped characterization of the sense in which reduction of a term M can be said to develop information out of better and better approximations of a normal form of M (if one exists). He extended the λ-calculus with a new constant Ω representing the undefined value, and defined an *approximate normal form* to be a term of the extended language (a λ-Ω-*term*) without β-redexes.[92] An approximate normal form A could then be said to match an ordinary λ-term except at occurrences of Ω in A, and this was exploited by Wadsworth in defining the set of *approximate normal forms* of M. One of his main results was the "limit theorem", [Wadsworth, 1976, Thm. 5.2], later called the "approximation theorem", [Barendregt, 1981, Thm. 19.2.2], which stated that the interpretation of a term M in Scott's D_∞ model is the least upper bound of the interpretations of its approximate normal forms.

A key technical concept throughout was that of a term M having a *head normal form*, that is, being convertible to a term of the form $\lambda x_1 \ldots x_n.zX_1 \ldots X_m$, for some $n, m \geq 0$. The identification of diverging computations with terms without a head normal form was found to be a natural one: for terms without head normal form, property (2) held, [Wadsworth, 1976, Cor. 5.5], and furthermore the interpretation of every such term in standard D_∞ models was the least element $\perp$ (ibid., Corollary 5.3).

Independently, Henk Barendregt, in June 1971 in his PhD thesis [Barendregt, 1971], proposed identifying diverging computations with what he called *unsolvable* terms, namely those λ-terms M such that

$$MX_1 \ldots X_k =_\beta \mathbf{I} \tag{3}$$

held for no sequence of terms $X_1, \ldots, X_k$ with $k \geq 0$. Barendregt was then attempting to construct a model for the λ-K-calculus based on recursion-theoretic

[91] Another important contribution of Wadsworth dating back from the same period (October 1970) was the idea of *continuations* as a technique for the denotational semantics of jumps. Publication of this, as a joint report [Strachey and Wadsworth, 1974], was delayed because of work on his thesis. The history of his work on continuations can be found in [Wadsworth, 2000]. The notion of continuation was also invented by several other people at about the same time; see the historical account in [Reynolds, 1993].

[92] This extension was suggested to Wadsworth by Dana Scott in conversation, in October 1970, see [Wadsworth, 1976, fn.4, p.506].

notions, related to the model of hereditarily recursive operations for the simply typed λ-calculus described in [Kreisel, 1958].[93]

A natural problem in this setting was to find out whether non-normalizing terms like $\Delta\Delta$ were defined in the recursion-theoretic interpretation.[94] However, the notion of undefinedness induced by divergence of partial recursive functions turned out to be too strong, and the identification of undefinedness with unsolvability emerged through an analysis of terms that do not have a normal form, yet possess a meaningful computational behavior that prevents them from all being identified consistently.[95] The consistency of the theory resulting from the identification of all unsolvable terms was then proved by proof-theoretic means in [Barendregt, 1971, Cor. 3.2.15 and Remark 3.2.16]. Also, by using unsolvable terms to represent undefined values $f(n)$ of a partial recursive function f, Barendregt obtained a simpler new proof that all partial recursive functions were λ-definable, [Barendregt, 1971, §1.3] or [Barendregt, 1981, §8.4].

It was Wadsworth (and Hyland, independently) who realized that a term M had no head normal form if and only if it was unsolvable [Wadsworth, 1971, p.94a], [Wadsworth, 1976, Cor. 5.3], allowing him to obtain a semantical version of Barendregt's consistency result by an immediate application of his limit theorem. (On the other hand, for the λI-calculus, Barendregt proved in 1973 that a term is unsolvable if and only if it has no normal form, [Barendregt, 1973].)

The interpretation of λ-terms in a D_∞ model induces a preorder on terms, defined by $M \leqslant N$ if and only if the interpretation of M is less than the interpretation of N for every choice of the interpretation of variables free in MN. The characterization of this preorder was given in [Wadsworth, 1976, Thm. 6.3] by introducing a notion of *semi-separability* of λ-terms, according to which M is semi-separable from N ($M \not\lesssim N$) iff there is a head context $C[\ ]$ such that $C[M] =_\beta \mathbf{I}$ but $C[N]$ is unsolvable. Then $M \not\lesssim N$ if and only if $M \not\leqslant N$. Independently, Hyland and Reiji Nakajima, then a student of Morris, obtained the same result for D_∞ [Hyland, 1975; Hyland, 1976; Nakajima, 1975], and Hyland found a similar characterization for the model $\mathcal{P}(\omega)$.[96] Behind all these studies lie, more or less explicitly, the techniques introduced by Böhm in the proof of his separability theorem [Böhm, 1968]: [Wadsworth, 1976] refers to a paper on "A general form of a theorem of Böhm and its application to Scott's models for the λ-calculus", that apparently was never published, while all the above results can be reformulated, following [Barendregt, 1977; Barendregt, 1981], using the notion of Böhm tree, cf. §7.1.

[93] [Barendregt, 1996] gives a very clear account of this attempt and of the positive results arising from it. These include, in particular, the partial validity of the ω-rule (i.e. if $FZ = GZ$ for all closed terms Z then $F = G$) for $\lambda\beta\eta$-conversion, when F and G are not *universal generators* like those constructed later in [Plotkin, 1974].

[94] In this interpretation, Δ becomes a partial recursive function φ_e such that $\varphi_e(x) \simeq \varphi_x(x)$.

[95] An example is the pair of terms $\lambda x.x(\Delta\Delta)\mathbf{K}$ and $\lambda x.x(\Delta\Delta)(\mathbf{KI})$, whose identification would lead immediately to the equation $\mathbf{K} = \mathbf{KI}$.

[96] Hyland played an important role in these developments: for example, in January 1972 he suggested the typed λ-calculus used in Wadsworth's proof of the approximation theorem, [Wadsworth, 1978]; see [Wadsworth, 1976, fn.5, p.508].

Jean-Jacques Lévy in his Thèse d'État [Lévy, 1978] used λ-Ω-terms as the building blocks of an "algebraic" model for the λ-calculus, much in the tradition of the French school of algebraic semantics of recursive programs [Nivat, 1973; Vuillemin, 1974b].

At about this time the question arose of saying what exactly the phrase "model of λ" meant in general. This was examined in 1978 by Hindley and Giuseppe Longo, [Hindley and Longo, 1980, §§2–3] and further in 1980 by Albert Meyer [Meyer, 1982, §6]. There turned out to be several variant concepts: λ-*model*, λ-*algebra*, *combinatory algebra*, as well as a category-based definition (see [Koymans, 1982, §§3–4]).[97] Each of these has given rise to further studies, some of which will be mentioned below. Another definition, related to λ as cylindric and polyadic algebras are related to first-order predicate calculus, was that of *lambda abstraction algebra*, introduced by Don Pigozzi and Antonio Salibra (and independently by Z. B. Diskin) around 1993; see [Pigozzi and Salibra, 1998] for a good overview.

We should mention here that the informal set theory in which Scott built D_∞ corresponded to the standard Zermelo-Fraenkel system, in which self-membership and infinite descending $\in$-chains $x_1 \ni x_2 \ni x_3 \ni \ldots$ are forbidden by the axiom of foundation. But "non-well-founded" set theories have been proposed at various times, in which such chains are allowed, and in such a theory the construction of a model of λ is very much simpler to carry out. This was first done by Michael von Rimscha in 1978, [Rimscha, 1980]. Some comments on non-well-founded models are in [Plotkin, 1993, pp.375–377], and [Aczel, 1988] is a general account of non-well-founded set theories.

9.3 Other models

After Scott's construction of the D_∞ models, Plotkin observed in 1972 [Plotkin, 1993, Part I] that a suitable generalization of the notion of graph of a function yielded a new family of models, now collectively called the *graph models*. For this generalization, he took an arbitrary atom ι and defined inductively a set T_C by the clauses:

- $\iota \in T_C$,
- if μ, ν are finite subsets of T_C, then $\langle \mu, \nu \rangle \in T_C$.

Then the powerset $\mathcal{P}(T_C)$ became a (non-extensional) model of the λ-calculus when application was defined by

$$x \cdot y \ =_{\text{def}} \ \bigcup \{\nu \mid \exists \mu \subseteq y.\ \langle \mu, \nu \rangle \in x\}. \tag{4}$$

A similar idea had also occurred to Bob Meyer in 1973–'74, stimulated by some work of Larry Powers, and resulted in a graph model of CL that he called the

[97] A combined account is in [Barendregt, 1981, 2nd. edition, Ch. 5].

Fool's model, [Meyer *et al.*, 1991, §5]. But this models reduction not equality. A related structure, for $\lambda\beta$-reduction, is described in [Plotkin, 1994, §4].

Other models in this family include those of [Scott, 1980a] and [Engeler, 1981], but the best-known of them is certainly $\mathcal{P}(\omega)$ of [Scott, 1976]. The model $\mathcal{P}(\omega)$ was motivated by "the project of making the connections with ordinary recursion theory easier to comprehend, since a satisfactory theory of computability and programming language semantics had to face this problem" [Scott, 1976, Appendix B]. The story of this model started at Easter 1973, when it was presented by Scott for the first time in Oberwolfach; later in the same year Scott realized that the idea behind it was essentially the same as Plotkin's definition (4), and that the same idea was "already implicit in a very precise form in much earlier work by Myhill-Shepherdson [Myhill and Shepherdson, 1955] and Friedberg-Rogers [Friedberg and Rogers, 1959] (see also [Rogers, 1967]) on *enumeration operators*" (Scott, ibid. p.576).[98]

The definition of application in $\mathcal{P}(\omega)$ needed the standard Cantor coding of pairs of natural numbers and an enumeration $\{e_n\}_{n\in\omega}$ of finite subsets of ω. Then the graph of a continuous function $f : \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ on the (complete algebraic) lattice $\mathcal{P}(\omega)$ was defined by

$$\{\langle n, m\rangle \mid n \in f(e_m)\},$$

and (4) became, under coding,

$$x \cdot y \ =_{\text{def}} \ \{n \in \omega \mid \exists e_m \subseteq y.\ \langle m, n\rangle \in x\}.$$

One of the aspects of $\mathcal{P}(\omega)$ that turned out to be important, especially for later investigations on the semantics of type theories, was the peculiar semantics of types allowed by the properties of this model. A *closure* is a continuous function $a : \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ such that $a(a(x)) = a(x) \supseteq x$ for all $x \in \mathcal{P}(\omega)$. Then a may be thought of as a type, namely the type of all $x \in \mathcal{P}(\omega)$ such that $a(x) = x$. For example, when a, b are closures, define $a \circ\!\!\to b$ as $\lambda u.\, b\circ u\circ a$, obtaining thus a closure that represents the type of (continuous) functions from a to b. Furthermore, there is a closure $\mathbf{V} : \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ such that $a \in \mathcal{P}(\omega)$ is a closure if and only if $\mathbf{V}(a) = a$. Such a $\mathbf{V}$ may be defined by $\mathbf{V} =_{\text{def}} \lambda a.\lambda x.\bigcup_{n=0}^{\infty} x^{(n)}$, where $x^{(0)} = \emptyset$ and $x^{(n+1)} = x \cup a(x^{(n)})$, or equivalently by:

$$\mathbf{V}(a)(x) \ =_{\text{def}} \ \bigcap\{y \in \mathcal{P}(\omega) \mid x \subseteq y \text{ and } a(y) \subseteq y\}.$$

The lattice of closures therefore directly yields a model for a type system with a type of all types, like Martin-Löf's first version of intuitionistic type theory. The construction of $\mathbf{V}$ is due essentially to Martin-Löf and Peter Hancock (circa 1974), and was devised in the course of Martin-Löf's attempts to extend Kreisel's topological interpretation to type theory. This extension was carried out essentially

[98] This story has had a recent – though implicit – continuation, in the study of *equilogical spaces* of [Scott, 1996], [Bauer *et al.*, 2004].

by means of spaces described by formal neighborhoods, similar to the information systems studied later by Scott as the canonical description of domains [Scott, 1982a]. In order to handle universes, they needed a "space of spaces", for which the quantifiers and the type operations were continuous. The construction of that space was described by Hancock (then a PhD student in Oxford) to Scott, and he soon adapted it to $\mathcal{P}(\omega)$.

Therefore, the domain $\mathcal{P}(\omega)$ was also the first example of a *universal domain*, in two different senses: every countably based algebraic (continuous) lattice is the image of a closure (retraction) over $\mathcal{P}(\omega)$, where a *retraction* is a continuous endofunction of $\mathcal{P}(\omega)$ idempotent with respect to composition, see [Scott, 1976]. This result set up a powerful tool for solving recursive domain equations (up to equality) inside $\mathcal{P}(\omega)$: in particular, the model D_∞ is represented by the least fixed-point of the closure $\lambda a.\, a \multimap a$ (see [Scott, 1976], and [Drakengren, 1996] who showed that this solution is unique, amending an incorrect statement in [Scott, 1976, p.553]).

These remarks gave rise to two different lines of research. On the one hand, there was a search for other similar universality results: progress in this area was based initially on modifications of the proofs for $\mathcal{P}(\omega)$, like Plotkin's universal domain $\mathbb{T}^\omega$ in [Plotkin, 1978b], or properties of the free Boolean algebra on $\aleph_0$ generators [Scott, 1982b]. Later it turned out that universal domains could be built uniformly by exploiting model-theoretic properties [Gunter, 1987; Gunter and Jung, 1990], in particular the amalgamation property [Droste and Göbel, 1993], that allowed homogeneous universal domains to be constructed for several classes of domains, unique up to isomorphism. On the other hand, the flexibility of retracts and related classes of endofunctions over universal domains in modelling type constructors suggested their use for interpreting higher-order type systems, especially the polymorphic λ-calculus. Types were interpreted as closures over $\mathcal{P}(\omega)$ in [McCracken, 1979], as retracts in [McCracken, 1982], and as finitary projections in [Amadio *et al.*, 1986]. Barendregt and Rezus also used closures for the semantics of Automath and related systems with dependent types [Barendregt and Rezus, 1983; Rezus, 1986].

10 DOMAIN THEORY

10.1 Classical domain theory

It is not completely clear what we can take as the starting date of a *theory* of domains, but by 1978 there was already a considerable bulk of results on categories of domains, many of them due to the efforts of Gordon Plotkin and collected in his notes for a lecture course in Pisa ([Plotkin, 1978a], complemented later by the notes [Plotkin, 1981]). The guiding analogy in that presentation was that between the complete partial order structure of domains and the approximation structure on domains induced by embedding-projection pairs. The iterative construction of the least fixed-point of a continuous endofunction of a domain corresponded in

this analogy to Scott's inverse limit construction; in the meantime the latter had been given a categorical description through the contributions of Reynolds, Wand, Smyth and Plotkin himself [Reynolds, 1972; Wand, 1974; Wand, 1977; Wand, 1979; Smyth and Plotkin, 1982].[99]

However, the main impetus to the study of the structure of domains was given not so much by the models of untyped λ as by the full abstraction problem for the typed λ-system PCF, the logic-free part of Milner's LCF ([Milner, 1972] and §8.1.2 above). This problem was described in 1975 in two papers, [Milner, 1977] and [Plotkin, 1977],[100] and concerns the relations between operational and denotational equivalence on programs.[101] Operationally, we say that two terms M and N of the same type σ are *observationally equivalent*, written $M \approx N$, when, for every context $C[\ \ ]$ with one hole of type σ such that $C[M]$ and $C[N]$ are programs, we have that $C[M] \Downarrow \mathbf{v}$ if and only if $C[N] \Downarrow \mathbf{v}$. This relation was first introduced in [Milner, 1975] in the context of a semantics of parallel computation, although it owed much to [Morris, 1968] and the early studies of the denotational properties of untyped λ-terms by Hyland, Wadsworth and Plotkin described in §9.2 above. Writing $[\![P]\!]$ for the denotation of a program P in some type structure, we say that programs P, Q are *equivalent* iff $[\![P]\!] = [\![Q]\!]$, and a denotational semantics is *fully abstract* [Milner, 1977] when $[\![P]\!] = [\![Q]\!]$ is equivalent to $P \approx Q$ for all PCF programs P, Q.

Plotkin showed that the standard model of PCF based on domains and Scott-continuous functions was not fully abstract [Plotkin, 1977, pp.234-236], but that it became so if one added parallel conditional operators $:\supset_\sigma$ of type $o \to (\sigma \to (\sigma \to \sigma))$, where $\sigma = o$ or $\sigma = \iota$. The interpretation of $:\supset_\sigma$ was the following continuous function:

$$:\supset_\sigma\ (p)(x)(y) \ = \ \begin{cases} x & \text{if } p = \mathbf{true} \\ y & \text{if } p = \mathbf{false} \\ x & \text{if } x = y \text{ and } p = \bot_o \\ \bot_\sigma & \text{if } x \neq y \text{ and } p = \bot_o. \end{cases}$$

On the other hand, [Milner, 1977, p.19, Cor. 3] proved that there is a unique fully abstract model for PCF (up to isomorphism); he constructed it essentially as a (quotiented) term model. The full abstraction problem consists basically in the

[99] Scott had already mentioned in [Scott, 1972, pp.128-129] a suggestion of Lawvere to the effect that the inverse limit construction of D_∞ could be carried out by regarding $[D \to D']$ as a functor (over the category of continuous lattices with projections as morphisms) whose second argument preserved inverse limits and whose first argument turned direct into inverse limits. The proof that $D_\infty \cong [D_\infty \to D_\infty]$ was then a direct consequence of these facts. A more general categorical framework for the solution of recursive domain equations was later described in [Freyd, 1990; Freyd, 1991; Freyd, 1992].

[100] See also the independent work carried out in the same year by Vladimir Yu. Sazonov and described in [Sazonov, 1976c; Sazonov, 1976b; Sazonov, 1976a].

[101] By a *program* in this context is meant a closed PCF-term of type either o, with *values* **true** and **false**, or ι, whose values are the numerals $\mathbf{n}$ for each natural number n. If a program M reduces to a value $\mathbf{v}$, we write $M \Downarrow \mathbf{v}$.

search for a syntax-independent description of Milner's model.[102]

The failure of full abstraction for the semantics of PCF based on Scott continuous functions is due to the presence of parallel functions, like the "*parallel or*" connective $\mathsf{V} : o \to (o \to o)$, already considered in [Platek, 1966, pp.127–131] and discussed there and in [Scott, 1969e], defined by the truth-table

V	$\bot$	**false**	**true**
$\bot$	$\bot$	$\bot$	**true**
false	$\bot$	**false**	**true**
true	**true**	**true**	**true**.

This fact drew attention to the problem of developing a semantic notion of *sequentiality* in order to "define a 'smaller' collection of domains containing only functions capable of deterministic realization" [Plotkin, 1977, p.236]. Definitions of sequential function had already been given in [Milner, 1977, p.20] and [Vuillemin, 1974a, §3.3]. However, these were coextensive only for Scott-continuous functions defined over flat domains (i. e., those where every chain had at most two elements) and depended upon a specific way of choosing the number of arguments of functions. A more general notion of sequentiality arose from the theory of *concrete domains* introduced in the Fall of 1975 by Gilles Kahn and Gordon Plotkin in [Kahn and Plotkin, 1978], where they axiomatized an abstract notion of "argument place" of a function by means of *information matrices* (or *concrete data structures*), over which computation proceeds by occurrence of *events* that consist in filling *cells* by *values*, subject to an *accessibility* relation that constrains the order in which cells are able to be filled. A domain is generated by an information matrix as the collection of its *configurations*, where each configuration is a set of events that represents the history of a computation over the matrix.

Though the general definition of sequential function based on concrete domains did not appear in [Kahn and Plotkin, 1978], the idea circulated widely, especially in the group of young French researchers at Sophia-Antipolis (including Gérard Berry and Pierre-Louis Curien). In [Berry, 1978], Berry used the algebraic model of [Lévy, 1976] to show that computation in pure untyped λ-calculus was essentially sequential in the sense of Kahn and Plotkin, extending Lévy's continuity theorem (these results are described in detail in [Barendregt, 1981, §§14.3–4]).

Later, Berry and Curien [Curien, 1979; Berry and Curien, 1982] introduced the notion of *sequential algorithm* over concrete data structures as a semantical framework for PCF. While sequential algorithms are not functions in the set-theoretic sense, they do form a cartesian closed category and therefore provide a

[102] We do not pursue further the technical aspects of full abstraction, although they determined much of the course of research on domains from 1975 onwards, for at least 20 years after. A comprehensive account of the strategies for solving it (and of the technical advances produced by those attempts) is in [Ong, 1995]. The solutions of the full abstraction problem obtained independently by Abramsky, Jagadeesan and Malacaria and Hyland and Ong, rely on a *game semantics* for PCF whose applications are still being investigated, [Abramsky *et al.*, 1994; Hyland and Ong, 2000]. See below, §10.3.

categorical model for PCF [Berry, 1981; Curien, 1986].[103] The calculations needed in the study of the category of concrete data structures and sequential algorithms also led Curien to introduce *categorical combinators* [Curien, 1985]; these became important in computational applications, starting from [Curien, 1986] (see also [Huet, 1990a]) who used them as instructions of an abstract machine [Cousineau *et al.*, 1987] for the implementation of the functional programming language ML.

A determinism condition on the computation of continuous functions that was close to sequentiality was devised by Berry, who called it *stability*, [Berry, 1976; Berry, 1978]. Stable functions satisfy a minimum data property whereby, for every finite approximation e of a value $f(x)$ there exists a minimum $d \sqsubseteq x$ such that $e \sqsubseteq f(d)$. Observe that the "parallel or" function is not stable. A more algebraic description of stability was possible in most cases of interest [Berry, 1976, Prop.II.2.1], just by saying that a Scott-continuous function is stable iff $f(x \wedge y) = f(x) \wedge f(y)$ provided x and y have a common upper bound. At higher orders, stable functions are ordered by $f \leq_{st} g$ iff $f(x) \sqsubseteq f(y) \wedge g(x)$ whenever $x \sqsubseteq y$; and with this definition, stable functions on appropriate domains (e.g. the dI-domains of [Berry, 1978, Def. 4.3.2], satisfying some of the properties of concrete domains) yield cartesian closed categories and, therefore, models of PCF.

The notion of stability was later rediscovered from a different point of view by [Girard, 1986]. Since the mid-1970s, Girard had been developing a theory of ordinals whose emphasis was on their geometric structure. The basic idea was that ordinal operations can be first defined on natural numbers and then extended to the whole class of ordinals by regarding it as a category whose objects are ordinals and whose morphisms are strictly increasing functions. The essential technical device to achieve this program was the notion of *dilator* [Girard, 1981], namely an endofunctor of the category of ordinals that preserves directed colimits and pullbacks. When specialised to domains regarded as categories, the preservation properties of dilators defined exactly the stable functions, and this was the starting point of the development of *qualitative domains* in [Girard, 1986] and [Girard, 1988, Annex A], and their simplification as *coherence spaces* in [Girard, 1986]. These yielded interesting models of his System $\mathbb{F}$ in which types were interpreted as domains. They also suggested the analysis of the (stable) function type $D \rightarrow_{st} E$ as $!D \multimap E$, where ! was a unary constructor on coherence spaces and $\multimap$ was *linear implication*, that would become the basic connective of *linear logic* [Girard, 1986], a logical system whose importance for the whole field of the semantics of computation is only beginning to be fully appreciated.

The representation of concrete domains as sets of configurations of information matrices was the central result in the monograph [Kahn and Plotkin, 1978, §7]. This suggested similar representation theorems for other categories of domains. In his lecture course at Merton College Oxford in 1981 [Scott, 1982b], Scott represented (Scott-) domains as the partially ordered sets of filters of structures that described the basis of a domain and therefore determined its whole partial or-

[103] There is a large literature on categorical models of typed and untyped λ-calculus, stimulated by the work of Lambek and also by [Scott, 1980b]: see [Koymans, 1982].

der structure (*neighbourhood systems*).[104] The next year, [Scott, 1982a] gave an equivalent representation of domains in terms of *information systems*: these were essentially abstract intuitionistic sequent calculi, which determined a set of *elements* (i.e., the elements of the represented domains) identified with Horn theories.

Another approach to the finitary representation of domains arose from an unexpected connection between Scott domains and the filter models built by means of intersection types (§8.4). The idea of using types to build models was not new, for example it was the main motivation of Plotkin's 1972 report [Plotkin, 1993, Part I, p.365].[105] Extending the ideas that led to the filter model of [Barendregt *et al.*, 1983], Coppo, Dezani, Honsell and Longo proved a precise correspondence between filter models built from what they called *extended applicative type structures* in [Coppo *et al.*, 1983] and [Coppo *et al.*, 1984], generalizing the set of intersection types, and the applicative information systems that form a subclass of the information systems used in [Scott, 1982a]. This was also recognized by Scott (ibid.), who remarked that intersection types could be regarded as just another representation of the finite elements of a domain. Filter models built from intersection types with additional constants – essentially a reformulation in this new context of [Coppo and Dezani, 1978] – were then used in [Coppo *et al.*, 1987] to build a model for untyped λ which was isomorphic to a non-standard D_∞ whose theory coincided with Morris' extensional theory [Morris, 1968].[106]

In the same years, Mike Smyth (and also Plotkin, in unpublished work) proposed looking at the open subsets of a topological space as computable properties of its points ([Smyth, 1983]; [Plotkin, 1981, Ch. 8, Exs.94–96]). Specialized to the Scott topology of a domain, this view complemented the identification of continuity and (abstract) computability suggested originally in [Scott, 1969e; Scott, 1969d] and [Plotkin, 1978a, §1]. This led to further representation results for domains, as characterized by the structure of the complete lattices of their open subsets. These results (thoroughly surveyed in [Abramsky and Jung, 1994]) were in effect an extension of the representation theory for Boolean algebras via Stone duality [Johnstone, 1982], influenced by closely related results in the theory of continuous lattices [Gierz *et al.*, 1980, §V.5] and the domain interpretation of Martin-Löf's type theory [Martin-Löf, 1983; Martin-Löf, 1986]. They were exploited in a logic of observable properties of domains first described systematically in [Vickers, 1989] and [Abramsky, 1991], and opened the way to the application of point-less topology in a computational setting, which was seen as very desirable from a constructive standpoint [Johnstone, 1982; Fourman and Grayson, 1982; Coquand, 1992, p.119],

[104] The *basis* of a domain D is the set of its finite elements, with the partial order inherited from D, where $d \in D$ is *finite* (or *algebraic*) iff $d \sqsubseteq \bigvee \Delta$ for some directed $\Delta \subseteq D$ implies $d \sqsubseteq e$ for some $e \in \Delta$. A cpo D is *algebraic* iff every $d \in D$ is the least upper bound of the directed set of finite elements below it, and a *Scott domain* is an algebraic cpo with a countable basis and least upper bounds of upper bounded subsets. This definition goes back to [Scott, 1969d].

[105] In fact, the graph models and filter models are presented together in [Plotkin, 1993, Part II, p.373ff.].

[106] The connections between filter models and the inverse limit construction have been explored in [Coppo *et al.*, 1984] and in [Plotkin, 1993].

and led to the point-free study of domain theory within Martin-Löf's intuitionistic type theory [Sambin, 1987].

10.2 Effective domains and Synthetic Domain Theory

We have remarked that computability considerations were at the basis of the very notions of domain and continuous function. Soon after those workers involved came to see algebraic and consistently complete partial orders as a convenient class of domains, much research was devoted to studying computability theory over *effectively given* domains; these allowed researchers to consider computable constructions in the concrete sense of classical recursion theory, [Constable and Egli, 1974], [Egli and Constable, 1976], [Smyth, 1977], [Kanda, 1979], [Plotkin, 1981, Ch. 7], [Scott, 1982b, Lecture VII]. Early ideas in this area had been put forward by Yu. Ershov using the theory of numbered sets proposed by Malcev, where a numbered set A consists of a set A with a surjective coding function e_A from A onto ω. In particular, in Ershov's framework the notion of *constructive domain* was defined, and it was proved that the Scott topology on such domains coincided with a topology that was induced in a natural way by its numbering (the *Malcev-Ershov topology*, see [Giannini and Longo, 1984; Rosolini, 1986] and the survey in [Longo, 1987]).

However, the resulting theory was still complicated by the need to manipulate the codes of elements of domains explicitly. A proposal to overcome these complications was put forward by Dana Scott, as the program of *synthetic domain theory*. A synthetic approach to computation consists in the study of constructive universes in which it is possible to regard domains just as special kinds of sets.[107] The name was suggested by an analogy with Synthetic Differential Geometry, where generalized manifolds are manipulated as sets of a special kind. Scott proposed such a research program in a talk at the Peripatetic Seminar on Sheaves and Logic in Sussex University in 1980, and later described its aim as that of having "a universe that combines logic and mathematics in a natural way and permits a development of finitary recursion theory in a sufficiently abstract way" [Scott, 1986].

The need for a treatment of partialness within intuitionistic universes of sets, described as toposes, led Scott's student Giuseppe Rosolini to introduce the basic notion in synthetic domain theory, namely that of *dominance*, a subset Σ of the set Ω of truth values that classifies the domains of definition of partial functions [Rosolini, 1986]. Intuitively, the elements of Σ are interpretations of the Σ^0_1 propositions in Ω. An important example of dominance arises in the effective topos $\mathcal{E}\mathit{ff}$ of [Hyland, 1982], where all functions on natural numbers are recursive, by setting

$$\Sigma \quad = \quad \{p \in \Omega \mid \ \exists f \in \mathbb{N}^{\mathbb{N}}.\ p \leftrightarrow (\exists z.\, f(z) = 0)\},$$

[107] Our short account of synthetic domain theory is based on the introduction to Rosolini's PhD thesis [Rosolini, 1986], on [Hyland, 1991], and on the section on synthetic domain theory by Moggi and Rosolini in [Jung (editor), 1996].

as was first done by Scott in a lecture in Pisa in the Fall of 1983. By exploiting this notion it was possible to define subcategories of the ambient topos whose objects might be taken as "Scott domains", and to prove for them the basic properties that categories of domains must have in order to support the denotational semantics of programming languages. The work on effective domains was therefore subsumed under the synthetic approach. Early contributions to this area include the work of Wesley Phoa [Phoa, 1990] on complete Σ-spaces as a synthetic counterpart to (pre)domains, and the alternative approach of [Freyd *et al.*, 1990]; and the axiomatic presentation of the fundamental ideas in [Hyland, 1991] and [Taylor, 1991]. Even more important, in demonstrating the unification achieved by means of a synthetic approach within the effective topos, was the discovery by Moggi and Hyland mentioned in §8.3 above that the category of modest sets yields a model of the polymorphic λ-calculus.

10.3 Game semantics

The full abstraction problem for PCF, that oriented much research on domain theory since the mid-70s [Milner, 1977; Plotkin, 1977], was given a solution by two independent teams: Abramsky, Jagadeesan and Malacaria from Imperial College, London [Abramsky *et al.*, 1994], and Hyland and Ong from Cambridge. Both groups announced their result in a message to the `types` mailing list on July 27th, 1993. Both used *games* as the semantic counterparts of types, and interpreted typed λ-terms as *strategies*, achieving what is called an *intensionally fully abstract* model for PCF, namely an algebraic model where every isolated element is definable.[108] The use of games (and strategies) did not come abruptly out of the blue, however, and has in fact a long history in the semantics of logical systems; under the correspondence between propositions and types, part of that tradition can be seen to be directly relevant to typed λ-calculi. From our point of view, the starting point of game semantics can be found in the works of Paul Lorenzen and his school from 1961 onward on dialogue games formalizing a debate between a Proponent and an Opponent, where the Opponent tries to attack a first-order formula defended by the Proponent [Lorenzen, 1961; Felscher, 1986; Lorenz, 2001]. The game-theoretical notion of validity was intended by Lorenzen to coincide with intuitionistic validity; a connection with typed λ-calculus was not made, even implicitly. This had to wait until the early 1990s, when it was exploited in the context of a semantical analysis of Girard's Linear Logic. The first mention of games in the context of linear logic goes back to a paper by Yves Lafont and Thomas Streicher [Lafont and Streicher, 1991], where a game was seen as a structure $\langle A^\star,\ A_\star,\ e : (A^\star \times A_\star) \to K\rangle$ consisting of two sets with a "payoff" function valued in a set K. A little earlier, Valeria de Paiva had given

[108] A quotient of the game model then yielded the order-extensional fully abstract model for PCF, as announced in a further message by Abramsky to the `types` mailing list on September 8th of the same year. It can be argued that the way to a direct (i.e., not quotiented) construction of the fully abstract model for PCF is barred by the undecidability of observational equivalence, shown by [Loader, 2001].

a related categorical account of Gödel's Dialectica interpretation [Paiva, 1989a; Paiva, 1989b] and had used it to interpret linear logic.[109]

The Lorenzen tradition was then brought into the semantics of linear logic by Andreas Blass in [Blass, 1992], building on his previous work on determinacy of (infinite) games [Blass, 1972]. With Blass' paper the use of dialogue games and strategies became a powerful tool for model construction, and in fact he was able to obtain a completeness theorem for the additive fragment of linear logic with respect to his games, stating that a sequent is provable if and only if there is a winning strategy for Proponent in the associated game.[110] One important drawback of this model was that composition of strategies was not associative, hence there was no category of games arising from Blass' work. On the other hand, a monoidal closed category of games (in the sense of [Conway, 1976]) had been already studied by André Joyal in [Joyal, 1977], who had noticed that it was a natural context for a "combinatorial" calculus of strategies. A category of games suitable for interpreting the multiplicative fragment of linear logic was introduced in 1992 by [Abramsky and Jagadeesan, 1994], who proved what they called a "full completeness" theorem: proofs in the multiplicative fragment of linear logic plus the MIX rule

$$\frac{\vdash \Gamma \quad \vdash \Delta}{\vdash \Gamma, \Delta}$$

correspond bijectively to winning strategies.[111] The perfect harmony of semantical objects (strategies) and syntactical structures (proofs) established by the full completeness theorem was exactly the kind of correspondence needed for building the fully abstract model of PCF of [Abramsky *et al.*, 1994], which in fact owed much to the framework set up for obtaining that theorem. In addition, the interpretation of composition of strategies as "parallel composition plus hiding" suggested by Abramsky [Abramsky, 1994; Abramsky and Jagadeesan, 1994] opened the way to the application of game semantics and ideas from linear logic to the theory of concurrent processes, especially process algebras; [Abramsky, 1994; Abramsky, 2000].[112]

Less closely related to the developments in linear logic was the approach of Hyland and Ong [Hyland and Ong, 2000]. They too used games, but their approach was more directly influenced by attempts to find a suitable notion of sequentiality

[109] Both these approaches are related to a method for building $*$-autonomous categories from suitably complete symmetric monoidal categories, invented by Po-Hsiang Chu and described in [Barr, 1979, Appendix]. [Seely, 1989] observed that $*$-autonomous categories yield categorical models for linear logic (see also [Barr, 1991]).

[110] Blass' model actually validates the Weakening rule (from $\Gamma \vdash B$ deduce $\Gamma, A \vdash B$), and is therefore a model of what is called *affine* logic.

[111] This kind of result can be traced back to Läuchli's completeness theorem for his interpretation of intuitionistic logic [Läuchli, 1965; Läuchli, 1970].

[112] The research in this direction was strongly influenced by Girard's idea of a *geometry of interaction* [Girard, 1987b; Girard, 1989a; Girard, 1989b], where the correctness criterion for (multiplicative) proof-nets – the natural deduction formulation of (the multiplicative fragment of) linear logic, [Girard, 1987a] – is expressed in terms of geometrical properties of graphs [Danos and Regnier, 1989].

at higher types, much in the same line as the theory of sequentiality of [Kahn and Plotkin, 1978] and [Berry and Curien, 1982], see [Ong, 1995], and by the compositional approach to games arising from [Blass, 1972] and [Joyal, 1977]. (In later presentations of game semantics, however, the technical framework set up by Hyland and Ong was used to build categories of games suitable for modelling (some fragment of) linear logic; see e.g., [Abramsky and McCusker, 1998].)

In the same direction, and with a bias towards the problem of extending Church's thesis to higher-type computability, there was already in the 1980s a series of papers by Kleene, who used dialogue games to interpret his higher-type recursion schemes [Kleene, 1978; Kleene, 1980; Kleene, 1982; Kleene, 1985]. Kleene's problem, the attempt to find "a class of functions which shall coincide with all the partial functions which are 'computable' or 'effectively decidable', so that Church's 1936 Thesis will apply with the higher types included" [Kleene, 1978, §1.2], led him to consider, though implicitly, the full abstraction problem for PCF.[113] Kleene's ideas were then pursued by Gandy and his student Giovanni Pani for continuous functionals [Gandy, 1993]. They emphasized the conditions of "no dangling question mark" for the game-theoretic analysis of computability in PCF, and this influenced, through informal discussions with Hyland, the formulation of the games of Hyland and Ong. Also seeming to be rooted in Kleene's work were the games used by [Nickau, 1994], that were in fact of the same kind as those of Hyland and Ong.

Beside the applications of game semantics to the interpretation of programming languages (surveyed, for example, in [Abramsky and McCusker, 1998]), there are developments of this area that pertain closely to the foundations of logic. An early example is [Coquand, 1995], where winning strategies in games associated to arithmetic propositions allow a reformulation of the analysis of their "finitist sense" provided by the consistency proofs of Gentzen and Novikoff. Here, for the first time, Lorenzen's view of proofs as strategies is applied to a dynamic analysis of cut-elimination. Another important example is *ludics*, a research program started by Girard in the late 1990s [Girard, 1998; Girard, 2000; Girard, 2001]. On a generically philosophical side, ludics is motivated by a critical revision of the traditional relations between syntax and semantics of logic, with a special emphasis on the meaning of logical rules: according to [Girard, 1998, p.215], this "is to be found in the well-hidden geometrical structure of the rules themselves: typically, negation should [be interpreted] by the exchange between *Player* and *Opponent*". The resulting emphasis on the symmetries of computation, already brought to the fore by linear logic, currently inspires much of the research at the border between logic and computation.

But this belongs to the 21st century, so our history stops here.

[113] Kleene's work was also independent from the parallel work on characterizations of sequentiality at higher-types by means of concrete data structures as in [Berry and Curien, 1982], that have been shown to have substantial game theoretic content, e.g., by François Lamarche [Lamarche, 1992] and by Pierre-Louis Curien [Curien, 1994], [Amadio and Curien, 1998, §14.3].

BIBLIOGRAPHY

[Abadi *et al.*, 1991] M. Abadi, L. Cardelli, P.-L. Curien, and J.-J. Lévy. Explicit substitutions. *Journal of Functional Programming*, 1:375–416, 1991.

[Abadi *et al.*, 1993] M. Abadi, L. Cardelli, and P.-L. Curien. Formal parametric polymorphism. *Theoretical Computer Science*, 121:9–58, 1993.

[Abdali, 1976] S. K. Abdali. An abstraction algorithm for combinatory logic. *Journal of Symbolic Logic*, 41:222–224, 1976.

[Abramsky and Jagadeesan, 1994] S. Abramsky and R. Jagadeesan. Games and full completeness for multiplicative linear logic. *Journal of Symbolic Logic*, 59:543–574, 1994.

[Abramsky and Jung, 1994] S. Abramsky and A. Jung. Domain theory. In S. Abramsky, D. Gabbay, and T. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 3, pages 1–168. Clarendon Press, Oxford, England, 1994.

[Abramsky and McCusker, 1998] S. Abramsky and G. McCusker. Game semantics. In U. Berger and H. Schwichtenberg, editors, *Computational Logic*, pages 1–56. Springer-Verlag, Berlin, 1998.

[Abramsky *et al.*, 1994] S. Abramsky, R. Jagadeesan, and P. Malacaria. Full abstraction for PCF (extended abstract). In H. Hagiya and J. C. Mitchell, editors, *Theoretical Aspects of Computer Software. International Symposium TACS '94*, volume 789 of *Lecture Notes in Computer Science*, pages 1–15. Springer-Verlag, Berlin, 1994.

[Abramsky, 1991] S. Abramsky. Domain theory in logical form. *Annals of Pure and Applied Logic*, 51:1–77, 1991.

[Abramsky, 1994] S. Abramsky. Proofs as processes. *Theoretical Computer Science*, 135:5–9, 1994.

[Abramsky, 2000] S. Abramsky. Process realizability. In F. L. Bauer and R. Steinbrüggen, editors, *Foundations of Secure Computation*, pages 167–180. IOS Press, Amsterdam, 2000.

[Aczel and Feferman, 1980] P. Aczel and S. Feferman. Consistency of the unrestricted abstraction principle using an intensional equivalence operator. In Hindley and Seldin [1980], pages 67–98.

[Aczel, 1980] P. Aczel. Frege structures and the notions of proposition, truth and set. In J. Barwise, H. J. Keisler, and K. Kunen, editors, *The Kleene Symposium*, pages 31–59. North-Holland Co., Amsterdam, 1980.

[Aczel, 1988] P. Aczel. *Non-Well-Founded Sets*. CSLI (Centre for the Study of Language and Information), Ventura Hall, Stanford Univ., Stanford, CA 94305-4115, U.S.A., 1988.

[Ajdukiewicz, 1935] K. Ajdukiewicz. Die syntaktische Konnexität. *Studia Philosophica*, 1:1–27, 1935. English transl: *Syntactic Connexion* in *Polish Logic 1920–1939*, ed. by Storrs McCall, Clarendon Press, Oxford, 1967, pp. 207–231.

[Amadio and Curien, 1998] R. Amadio and P.-L. Curien. *Domains and Lambda-Calculi*, volume 46 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, England, 1998.

[Amadio *et al.*, 1986] R. Amadio, K. Bruce, and G. Longo. The finitary projection model for second order lambda calculus and solutions to higher order domain equations. In *Proceedings First Annual IEEE Symposium on Logic In Computer Science*, pages 122–130. IEEE Computer Society, Los Alamitos, California, U.S.A., 1986.

[Anderson, 1998] C. A. Anderson. Alonzo Church's contributions to philosophy and intensional logic. *Bulletin of Symbolic Logic*, 4:129–171, 1998.

[Andrews, 1965] P. B. Andrews. *A Transfinite Type Theory with Type Variables*. North-Holland Co., Amsterdam, 1965.

[Andrews, 1971] P. B. Andrews. Resolution in type theory. *Journal of Symbolic Logic*, 36:414–432, 1971.

[Asperti and Longo, 1991] A. Asperti and G. Longo. *Categories, Types and Structures. An Introduction to Category Theory for the Working Computer Scientist*. Foundations of Computing. M.I.T. Press, Cambridge, Mass., U.S.A., 1991.

[Asperti, 1995] A. Asperti. Linear logic, comonads and optimal reductions. *Fundamenta Informaticae*, 22(1–2):3–22, 1995.

[Backus, 1978] J. Backus. Can programming be liberated from the von Neumann style? *Communications of the ACM*, 21(8):613–641, 1978.

[Bar-Hillel, 1953] Y. Bar-Hillel. A quasi-arithmetical notation for syntactic description. *Language*, 29:47–58, 1953.

[Bar-Hillel, 1959] Y. Bar-Hillel. Decision procedures for structure in natural languages. *Logique et analyse*, 2:19–29, 1959.

[Barendregt and Rezus, 1983] H. P. Barendregt and A. Rezus. Semantics for classical AUTOMATH and related systems. *Information and Control*, 59:127–147, 1983. (Journal now called *Information and Computation*).

[Barendregt *et al.*, 1976] H. P. Barendregt, J. Bergstra, J. W. Klop, and H. Volken. Representability in lambda algebras. *Koninkl. Nederlands Akad. van Wetensch. Proc. Ser. A*, 79:377–387, 1976. Vol. also publ. as *Indagationes Mathematicae* 38.

[Barendregt *et al.*, 1983] H. P. Barendregt, M. Coppo, and M. Dezani. A filter lambda model and the completeness of type assignment. *Journal of Symbolic Logic*, 48:931–940, 1983.

[Barendregt *et al.*, 1993] H. Barendregt, M. Bunder, and W. Dekkers. Systems of illative combinatory logic complete for first order propositional and predicate calculus. *Journal of Symbolic Logic*, 58:769–788, 1993.

[Barendregt, 1971] H. P. Barendregt. *Some extensional term models for combinatory logics and λ-calculi*. PhD thesis, Univ. Utrecht, Netherlands, 1971.

[Barendregt, 1973] H. P. Barendregt. A characterization of terms of the λ-I-calculus having a normal form. *Journal of Symbolic Logic*, 38:441–445, 1973.

[Barendregt, 1976] H. P. Barendregt. A global representation of the recursive functions in the λ-calculus. *Theoretical Computer Science*, 3:225–242, 1976.

[Barendregt, 1977] H. P. Barendregt. The type-free lambda calculus. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 1091–1132. North-Holland Co., Amsterdam, 1977.

[Barendregt, 1981] H. P. Barendregt. *The Lambda Calculus, its Syntax and Semantics*. North-Holland Co., Amsterdam, 1981. 2nd (revised) edn. 1984.

[Barendregt, 1993] H. P. Barendregt. Constructive proofs of the range property in lambda calculus. *Theoretical Computer Science*, 121 (Böhm Volume):59–69, 1993.

[Barendregt, 1996] H. P. Barendregt. Kreisel, lambda calculus, a windmill and a castle. In P. G. Odifreddi, editor, *Kreiseliana, about and around Georg Kreisel*, pages 3–14. A. K. Peters, Wellesley, Massachusetts, 1996.

[Barendregt, 1997] H. P. Barendregt. The impact of the lambda calculus in logic and computer science. *Bulletin of Symbolic Logic*, 3:181–215, 1997.

[Barr, 1979] M. Barr. *$*$-autonomous Categories*, volume 752 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1979.

[Barr, 1991] M. Barr. $*$-autonomous categories and linear logic. *Mathematical Structures in Computer Science*, 1:159–178, 1991.

[Bates and Constable, 1985] J. L. Bates and R. L. Constable. Proofs as programs. *ACM Transactions on programming Languages and Systems*, 7(1):113–136, 1985.

[Bauer *et al.*, 2004] A. Bauer, L. Birkedal, and D. S. Scott. Equilogical spaces. *Theoretical Computer Science*, 315:35–59, 2004.

[Ben-Yelles, 1979] C-B. Ben-Yelles. *Type-assignment in the Lambda-calculus*. PhD thesis, Math. Dept., Univ. Wales Swansea, Swansea SA2 8PP, U.K., 1979.

[Bénabou, 1985] J. Bénabou. Fibered categories and the foundations of naive category theory. *Journal of Symbolic Logic*, 50:10–37, 1985.

[Benaissa *et al.*, 1996] Z. Benaissa, D. Briaud, P. Lescanne, and J. Rouyer-Degli. λv, a calculus of explicit substitutions which preserves strong normalisation. *Journal of Functional Programming*, 6:699–722, 1996.

[Bergstra and Klop, 1980] J. Bergstra and J. W. Klop. Invertible terms in the lambda calculus. *Theoretical Computer Science*, 11:19–37, 1980.

[Bernays and Schönfinkel, 1928] P. Bernays and M. Schönfinkel. Zum Entscheidungsproblem der mathematischen Logik. *Mathematische Annalen*, 99:342–372, 1928.

[Bernays, 1958] P. Bernays. Über eine natürliche Erweiterung des Relationenkalkuls. In A. Heyting, editor, *Constructivity in Mathematics*, pages 1–14. North-Holland Co., Amsterdam, 1958.

[Berry and Curien, 1982] G. Berry and P.-L. Curien. Sequential algorithms on concrete data structures. *Theoretical Computer Science*, 20:265–321, 1982.

[Berry, 1976] G. Berry. Bottom-up computation of recursive programs. *RAIRO Informatique Théorique et Applications*, 10:47–82, 1976.

[Berry, 1978] G. Berry. Stable models of typed λ-calculi. In G. Ausiello and C. Böhm, editors, *Automata, Languages and Programming, Fifth Colloquium*, volume 62 of *Lecture Notes in Computer Science*, pages 72–89. Springer-Verlag, Berlin, 1978.

[Berry, 1981] G. Berry. On the definition of λ-calculus models. In J. Diaz and I. Ramos, editors, *International Colloquium on Formalization of Programming Concepts, Proceedings*, volume 107 of *Lecture Notes in Computer Science*, pages 218–230. Springer-Verlag, Berlin, 1981.

[Bethke, 2000] I. Bethke. Annotated Bibliography of Lambda Calculi, Combinatory Logics and Type Theory. URL www.adam.science.uva.nl/ inge/, 2000. File type '.ps', 587 pp. if printed.

[Blass, 1972] A. Blass. Degrees of indeterminacy of games. *Fundamenta Mathematicae*, 77:151–166, 1972.

[Blass, 1992] A. Blass. A game semantics for linear logic. *Annals of Pure and Applied Logic*, 56(1–3):183–220, 1992.

[Bloo and Geuvers, 1999] R. Bloo and J. H. Geuvers. Explicit substitution: on the edge of strong normalization. *Theoretical Computer Science*, 211:375–395, 1999.

[Böhm and Berarducci, 1985] C. Böhm and A. Berarducci. Automatic synthesis of typed λ-programs on term algebras. *Theoretical Computer Science*, 39:135–154, 1985.

[Böhm and Dezani, 1972] C. Böhm and M. Dezani. A CUCH machine: The automatic treatment of bound variables. *International Journal of Computer and Information Sciences*, 1:171–191, 1972. See also *Notes on "A CUCH machine: the automatic treatment of bound variables"*, *ibid* 2 (1973), 157–160.

[Böhm and Dezani, 1974] C. Böhm and M. Dezani. A parenthesis machine for string manipulation. *Revue Français d'Automatique, Informatique et Recherche Opérationelle*, 8:37–46, 1974.

[Böhm and Dezani, 1975] C. Böhm and M. Dezani. λ-terms as total or partial functions on normal forms. In Böhm [1975], pages 96–121.

[Böhm and Dezani, 1989] C. Böhm and M. Dezani. Combinatory logic as monoids. *Fundamenta Informaticae*, 12(4):525–540, 1989.

[Böhm and Giovannucci, 1964] C. Böhm and R. Giovannucci. Circuiti sequenziali e analogici e loro descrizione mediante il CUCH. Istituto per le Applicazioni del Calcolo, C.N.R., Roma, 1964.

[Böhm and Gross, 1966] C. Böhm and W. Gross. Introduction to the CUCH. In E. Caianiello, editor, *Automata Theory*, pages 35–65. Academic Press, New York, 1966. (Preprints available 1964).

[Böhm *et al.*, 1977] C. Böhm, M. Coppo, and M. Dezani. Termination tests inside λ-calculus. In A. Salomaa and M. Steinby, editors, *Automata, Languages and Programming, Fourth Colloquium*, volume 52 of *Lecture Notes in Computer Science*, pages 95–110. Springer-Verlag, Berlin, 1977.

[Böhm, 1954] C. Böhm. Calculatrices digitales. Du déchiffrage des formules logico-mathématiques par la machine même dans la conception du programme. *Annali di Matematica Pura e Applicata*, 37(4):1–51, 1954.

[Böhm, 1966] C. Böhm. The CUCH as a formal and description language. In T. B. Steel, Jr., editor, *Formal Language Description Languages for Computer Programming*, pages 179–197. North-Holland Co., Amsterdam, 1966. Proc. conference in 1964.

[Böhm, 1968] C. Böhm. Alcune proprietà delle forme β-η-normali nel λ-K-calcolo. Pubblicazione no. 696, Istituto per le Applicazioni del Calcolo, C.N.R., Roma, 1968.

[Böhm, 1975] C. Böhm, editor. *λ-Calculus and Computer Science Theory*, volume 37 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1975.

[Böhm, 1979] C. Böhm. Un modèle arithmétique des termes de la logique combinatoire. In B. Robinet, editor, *Lambda Calcul et Sémantique Formelle des Langages de Programmation. Actes de la Sixième Ecole de Printemps d'Informatique Theorique, La Châtre, 1978*, pages 97–108. LITP et ENSTA, Paris, 1979.

[Böhm, 1980] C. Böhm. Logic and computers. In E. Agazzi, editor, *Modern Logic — A Survey*, pages 297–309. D. Reidel Co., Dordrecht, Netherlands, 1980.

[Böhm, 1982] C. Böhm. Combinatory foundations of functional programming. In *Proceedings of the 1982 ACM Symposium on Lisp and Functional Programming*, pages 29–36. Association for Computing Machinery, New York, 1982.

[Böhm, 1986] C. Böhm. Reducing recursion to iteration by algebraic extension. In B. Robinet and R. Wilhelm, editors, *European Symposium on Programming*, volume 213 of *Lecture Notes in Computer Science*, pages 111–118. Springer-Verlag, Berlin, 1986.

[Böhm, 1988a] C. Böhm. Functional programming and combinatory algebras. In M. P. Chytil, L. Janiga, and V. Koubek, editors, *Mathematical Foundations of Computer Science*, volume 324 of *Lecture Notes in Computer Science*, pages 14–26. Springer-Verlag, Berlin, 1988.

[Böhm, 1988b] C. Böhm. Reducing recursion to iteration by means of pairs and n-tuples. In M. Boscarol, L. Carlucci Aiello, and G. Levi, editors, *Foundations of Logic and Functional Programming*, volume 306 of *Lecture Notes in Computer Science*, pages 58–66. Springer-Verlag, Berlin, 1988. Proc. workshop in 1986.

[Böhm, 1989] C. Böhm. Subduing self-application. In G. Ausiello, M. Dezani, and S. Ronchi della Rocca, editors, *Automata, Languages and Programming, 16th International Colloquium*, volume 372 of *Lecture Notes in Computer Science*, pages 108–122. Springer-Verlag, Berlin, 1989.

[Broda and Damas, 1999] S. Broda and L Damas. Counting a type's principal inhabitants. In J.-Y. Girard, editor, *Typed Lambda Calculi and Applications, 4th International Conference, TLCA'99*, volume 1581 of *Lecture Notes in Computer Science*, pages 69–82. Springer-Verlag, Berlin, 1999. Fuller version: *Fundamenta Informaticae* 45 (2001), 33–51.

[Bruce *et al.*, 1990] K. Bruce, A. A. Meyer, and J. C. Mitchell. The semantics of second-order lambda calculus. In Huet [1990b], pages 273–284.

[Bruijn, 1970] N. G. de Bruijn. The mathematical language AUTOMATH. In M. Laudet, D. Lacombe, L. Nolin, and M. Schützenberger, editors, *Symposium on Automatic Demonstration*, volume 125 of *Lecture Notes in Mathematics*, pages 29–61. Springer-Verlag, Berlin, 1970.

[Bruijn, 1972] N. G. de Bruijn. Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation. *Indagationes Mathematicae*, 34:381–392, 1972.

[Bruijn, 1980] N. G. de Bruijn. A survey of the project AUTOMATH. In Hindley and Seldin [1980], pages 579–606.

[Bruijn, 1994] N. G. de Bruijn. Reflections on Automath. In Nederpelt et al. [1994], pages 201–228.

[Bunder and Dekkers, 2005] M. W. Bunder and W. Dekkers. Equivalences between pure type systems and systems of illative combinatory logic. *Notre Dame Journal of Formal Logic*, 46:181–205, 2005.

[Bunder, 1983a] M. W. Bunder. A one axiom set theory based on higher order predicate calculus. *Archiv für Mathematische Logik*, 23:99–107, 1983.

[Bunder, 1983b] M. W. Bunder. Predicate calculus of arbitrarily high finite order. *Archiv für Mathematische Logik*, 23:1–10, 1983.

[Bunder, 1983c] M. W. Bunder. Set theory in predicate calculus with equality. *Archiv für Mathematische Logik*, 23:109–113, 1983.

[Burali-Forti, 1894] C. Burali-Forti. *Logica Matematica*. Manuali Hoepli. U. Hoepli, Milano, 1894.

[Burge, 1978] W. H. Burge. *Recursive Programming Techniques*. Addison-Wesley, Reading, Mass., U.S.A., 1978.

[Burstall, 1977] R. M. Burstall. Design considerations for a functional programming language. In *The Software Revolution; Proceedings of the Infotech State of the Art Conference, Copenhagen 1977*, pages 45–57. Infotech and Pergamon Press, U.K., 1977.

[Cardelli and Wegner, 1985] L. Cardelli and P. Wegner. On understanding types, data abstraction and polymorphism. *ACM Computing Surveys*, 17:471–522, 1985.

[Cardone and Coppo, 1991] F. Cardone and M. Coppo. Type inference with recursive types. Syntax and Semantics. *Information and Computation*, 92(1):48–80, 1991.

[Chauvin, 1979] A. Chauvin. Theory of objects and set theory: introduction and semantics. *Notre Dame Journal of Formal Logic*, 10:37–54, 1979. (English transl. of part of thesis *Théorie des Objets et Théorie des Ensembles*, Université de Clermont-Ferrand, France, 1974.).

[Cheatham Jr. *et al.*, 1968] T. E. Cheatham Jr., P. Fischer, and P. Jorrand. On the basis of ELF – An extensible language facility. In *AFIPS Conference Proceedings, San Francisco, California*, pages 937–948. Thompson Book Co., Washington, D.C., 1968.

[Church and Rosser, 1936] A. Church and J. B. Rosser. Some properties of conversion. *Transactions of the American Mathematical Society*, 39:472–482, 1936.

[Church, 1932] A. Church. A set of postulates for the foundation of logic. *Annals of Mathematics, Series 2*, 33:346–366, 1932.

[Church, 1933] A. Church. A set of postulates for the foundation of logic (second paper). *Annals of Mathematics, Series 2*, 34:839–864, 1933.

[Church, 1935] A. Church. A proof of freedom from contradiction. *Proceedings of the National Academy of Sciences of the U.S.A.*, 21:275–281, 1935.

[Church, 1936a] A. Church. A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1:40–41, 1936. See also correction in pp. 101–102.

[Church, 1936b] A. Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58:345–363, 1936.

[Church, 1937] A. Church. Combinatory logic as a semi-group. *Bulletin of the American Mathematical Society*, 43:333, 1937. Only an abstract.

[Church, 1940] A. Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5:56–68, 1940.

[Church, 1941] A. Church. *The Calculi of Lambda Conversion*. Princeton Univ. Press, Princeton, New Jersey, U.S.A., 1941. Reprinted 1951 and 2000.

[Church, 1951a] A. Church. A formulation of the logic of sense and denotation. In P. Henle, H. Kallen, and S. Langer, editors, *Structure, Method and Meaning, Essays in Honor of Henry M. Sheffer*, pages 3–24. Liberal Arts Press, New York, 1951. See also abstract with same title, in J. Symbolic Logic 11 (1946), p. 31.

[Church, 1951b] A. Church. The weak theory of implication. In A. Menne, A. Wilhelmy, and H. Angstl, editors, *Kontrolliertes Denken, Untersuchungen zum Logikkalkül und zur Logik der Einzelwissenschaften (Festgabe zum 60. Geburtstag von Prof. W. Britzelmayr)*, pages 22–37. Verlag Karl Alber, München, 1951. (Sonderdruck).

[Church, 1956] A. Church. *Introduction to Mathematical Logic*. Princeton Univ. Press, Princeton, New Jersey, U.S.A., 1956.

[Church, 1964] A. Church, 7 July 1964. Unpublished letter to Harald Dickson.

[Church, 1973] A. Church. Outline of a revised formulation of the logic of sense and denotation, Part I. *Noûs*, 7:24–33, 1973.

[Chwistek, 1922] L. Chwistek. Über die Antinomien der Prinzipen der Mathematik. *Mathematische Zeitschrift*, 14:236–243, 1922.

[Constable and Egli, 1974] R. L. Constable and H. Egli. Computability on continuous higher types and its role in the semantics of programming languages. Technical report TR-74-209, Department of Computer Science, Cornell Univ. , Ithaca, New York, 1974.

[Constable and O'Donnell, 1978] R. L. Constable and M. J. O'Donnell. *A Programming Logic*. Winthrop, Cambridge, 1978.

[Constable and others, 1986] R. L. Constable et al. *Implementing Mathematics with the Nuprl Proof Development System*. Prentice-Hall, Englewood Cliffs, N.J., U.S.A., 1986.

[Constable, 1971] R. L. Constable. Constructive mathematics and automatic program writers. In *Proceedings of the IFIP Congress*, pages 229–233, Ljubljana, 1971.

[Constable, 1980] R. L. Constable. Programs and types. In *21st IEEE Symposium on the Foundations of Computer Science*, pages 118–128, Los Alamitos, California, U.S.A., 1980. IEEE Computer Society.

[Constable, 1982] R. L. Constable. Programs as proofs. Technical report TR-82-532, Dept. of Computer Science, Cornell Univ. , Ithaca, New York, 1982.

[Constable, 1985] R. L. Constable. Constructive mathematics as a programming logic I: Some principles of theory. *Annals of Discrete Mathematics*, 24:21–38, 1985.

[Conway, 1976] J. H. Conway. *On Numbers and Games*, volume 6 of *London Mathematical Society Monographs*. Academic Press, London, 1976.

[Coppo and Dezani, 1978] M. Coppo and M. Dezani. A new type assignment for λ-terms. *Archiv für Mathematische Logik*, 19:139–156, 1978.

[Coppo and Ferrari, 1993] M. Coppo and A. Ferrari. Type inference, abstract interpretation and strictness analysis. *Theoretical Computer Science*, 121:113–143, 1993.

[Coppo *et al.*, 1979] M. Coppo, M. Dezani, and P. Sallé. Functional characterization of some semantic equalities inside λ-calculus. In H. Maurer, editor, *Automata, Languages and Programming, Sixth Colloquium*, volume 71 of *Lecture Notes in Computer Science*, pages 133–146. Springer-Verlag, Berlin, 1979.

[Coppo *et al.*, 1980] M. Coppo, M. Dezani, and B. Venneri. Principal type-schemes and λ-calculus semantics. In Hindley and Seldin [1980], pages 535–560.

[Coppo *et al.*, 1981] M. Coppo, M. Dezani, and B. Venneri. Functional characters of solvable terms. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 27:45–58, 1981.

[Coppo *et al.*, 1983] M. Coppo, M. Dezani, F. Honsell, and G. Longo. Applicative information systems. In G. Ausiello and M. Protasi, editors, *Colloquium on Trees and Algebra in Programming*, volume 159 of *Lecture Notes in Computer Science*, pages 33–64. Springer-Verlag, Berlin, 1983.

[Coppo *et al.*, 1984] M. Coppo, M. Dezani, F. Honsell, and G. Longo. Extended type structures and filter lambda models. In G. Lolli, G. Longo, and A. Marcja, editors, *Logic Colloquium '82*, pages 241–262. North-Holland Co., Amsterdam, 1984.

[Coppo *et al.*, 1987] M. Coppo, M. Dezani, and M. Zacchi. Type theories, normal forms and D_∞-lambda models. *Information and Computation*, 72:85–116, 1987.

[Coppo, 1985] M. Coppo. A completeness theorem for recursively defined types. In W. Brauer, editor, *Automata, Languages and Programming, 12th International Colloquium*, volume 194 of *Lecture Notes in Computer Science*, pages 120–129, Berlin, 1985. Springer-Verlag.

[Coquand and Dybjer, 1994] T. Coquand and P. Dybjer. Inductive definitions and type theory: an introduction (preliminary version). In P. S. Thiagarajan, editor, *Foundations of Software Technology and Theoretical Computer Science, 14th Conference, Madras, India, December 15-17, 1994, Proceedings*, pages 60–76, 1994.

[Coquand and Huet, 1985] T. Coquand and G. Huet. Constructions: a higher order proof system for mechanizing mathematics. In B. Buchberger, editor, *EUROCAL'85, Proceedings Volume 1*, volume 203 of *Lecture Notes in Computer Science*, pages 151–184, Berlin, 1985. Springer-Verlag.

[Coquand and Huet, 1988] T. Coquand and G. Huet. The calculus of constructions. *Information and Computation*, 76:95–120, 1988.

[Coquand, 1985] T. Coquand. Une théorie des constructions. Thèse de Troisième Cycle, Université Paris VII, 1985.

[Coquand, 1986] T. Coquand. An analysis of Girard's paradox. In *Proceedings First Annual IEEE Symposium on Logic In Computer Science*, pages 227–236. IEEE Computer Society, Los Alamitos, California, U.S.A., 1986.

[Coquand, 1990] T. Coquand. Metamathematical investigations of a calculus of constructions. In P. Odifreddi, editor, *Logic and Computer Science*, volume 31 of *APIC Studies in Data Processing*, pages 91–122. Academic Press, London, 1990.

[Coquand, 1992] T. Coquand. An intuitionistic proof of Tychonoff's theorem. *Journal of Symbolic Logic*, 57:28–32, 1992.

[Coquand, 1995] T. Coquand. A semantics of evidence for classical arithmetic. *Journal of Symbolic Logic*, 60(1):325–337, 1995.

[Cousineau *et al.*, 1987] G. Cousineau, P.-L. Curien, and M. Mauny. The categorical abstract machine. *Science of Computer Programming*, 8:173–202, 1987.

[Craig, 1974] W. Craig. *Logic in an Algebraic Form, Three Languages and Theories*. North-Holland Co., Amsterdam, 1974.

[Crossley, 1975] J. N. Crossley. Reminiscences of Logicians. In J. Crossley, editor, *Algebra and Logic*, volume 450 of *Lecture Notes in Mathematics*, pages 1–62. Springer-Verlag, Berlin, 1975.

[Curien *et al.*, 1996] P.-L Curien, T. Hardin, and J.-J. Lévy. Confluence properties of weak and strong calculi of explicit substitutions. *Journal of the Association for Computing Machinery*, 43:362–397, 1996.

[Curien, 1979] P.-L. Curien. Algorithmes sequentiels sur structures de donnees concretes. In B. Robinet, editor, *Lambda Calcul et Sémantique Formelle des Langages de Programmation. Actes de la Sixième Ecole de Printemps d'Informatique Théorique*, pages 157–182. LITP et Ecole Nationale Supérieure de Techniques Avancées, 1979.

[Curien, 1985] P.-L. Curien. Typed categorical combinatory logic. In H. Ehrig, C. Floyd, M. Nivat, and J. Thatcher, editors, *Mathematical Foundations of Software Development, Proceedings 1985, Volume 1, CAAP'85*, volume 185 of *Lecture Notes in Computer Science*, pages 157–172. Springer-Verlag, Berlin, 1985.

[Curien, 1986] P.-L. Curien. *Categorical Combinators, Sequential Algorithms and Functional Programming*. Pitman (London) and Wiley (New York), 1986. (2nd edn. 1993 publ. by Birkhäuser, U.S.A.).

[Curien, 1994] P.-L. Curien. On the symmetry of sequentiality. In S. Brookes, M. Main, A. Melton, M. Mislove, and D. Schmidt, editors, *Mathematical Foundations of Programming Semantics, 9th International Conference*, volume 802 of *Lecture Notes in Computer Science*, pages 29–71. Springer-Verlag, Berlin, 1994.

[Curry and Feys, 1958] H. B. Curry and R. Feys. *Combinatory Logic, Volume I*. North-Holland Co., Amsterdam, 1958. (3rd edn. 1974).

[Curry *et al.*, 1972] H. B. Curry, J. R. Hindley, and J. P. Seldin. *Combinatory Logic, Volume II*. North-Holland Co., Amsterdam, 1972.

[Curry, 1927] H. B. Curry. *Schönfinkel, M. Über die Bausteine der mathematischen Logik*, 1927. Note dated Nov. 28th. 1927, 3 pages, filed as T271128A in Curry archive, Pennsylvania State Univ., U.S.A. Reproduced on website of Southern Alberta Digital Library, http://www.sadl.uleth.ca.

[Curry, 1930] H. B. Curry. Grundlagen der kombinatorischen Logik. *American Journal of Mathematics*, 52:509–536, 789–834, 1930.

[Curry, 1932] H. B. Curry. Some additions to the theory of combinators. *American Journal of Mathematics*, 54:551–558, 1932.

[Curry, 1934a] H. B. Curry. Functionality in combinatory logic. *Proceedings of the National Academy of Sciences of the U.S.A.*, 20:584–590, 1934.

[Curry, 1934b] H. B. Curry. Some properties of equality and implication in combinatory logic. *Annals of Mathematics, Series 2*, 35:849–860, 1934.

[Curry, 1936] H. B. Curry. First properties of functionality in combinatory logic. *Tôhoku Mathematical Journal*, 41:371–401, 1936.

[Curry, 1937] H. B. Curry. Review of Church's lecture notes *Mathematical Logic* given at Princeton Univ. in 1936. *Journal of Symbolic Logic*, 2:39–40, 1937.

[Curry, 1942a] H. B. Curry. The combinatory foundations of mathematical logic. *Journal of Symbolic Logic*, 7:49–64, 1942.

[Curry, 1942b] H. B. Curry. The inconsistency of certain formal logics. *Journal of Symbolic Logic*, 7:115–117, 1942.

[Curry, 1948] H. B. Curry. The logical structure of grammar, 1948. Unfinished manuscript mimeographed at Univ. Chicago and privately distributed, from a lecture given in Nov. 1948.

[Curry, 1949] H. B. Curry. A simplification of the theory of combinators. *Synthese*, 7:391–399, 1949.

[Curry, 1951] H. B. Curry. *Outlines of a Formalist Philosophy of Mathematics*. North-Holland Co., Amsterdam, 1951.

[Curry, 1954] H. B. Curry. The logic of program composition. In *Applications Scientifiques de la Logique Mathématique, Actes du Deuxieme Colloque International de Logique Mathématique, Paris 1952*, pages 97–102. Gauthier-Villars, Paris, 1954.

[Curry, 1961] H. B. Curry. Some logical aspects of grammatical structure. In R. Jakobson, editor, *Structure of Language and its Mathematical Aspects*, number 12 in Proceedings of Symposia in Applied Mathematics, pages 56–68. American Mathematical Society, Providence, R.I., U.S.A., 1961.

[Curry, 1963] H. B. Curry. *Foundations of Mathematical Logic*. McGraw-Hill, New York, 1963. Reprinted 1977 by Dover, Inc., New York.

[Curry, 1964] H. B. Curry. The elimination of variables by regular combinators. In M. Bunge, editor, *The Critical Approach to Science and Philosophy*, pages 127–143. Free Press, U.S.A., 1964.

[Curry, 1966] H. B. Curry. Technique for evaluating principal functional character, 1966. Note dated March 17th. 1966, 5 pages, filed as T660317A in Curry archive, Pennsylvania State Univ. , U.S.A.

[Curry, 1969a] H. B. Curry. Modified basic functionality in combinatory logic. *Dialectica*, 23:83–92, 1969.

[Curry, 1969b] H. B. Curry. The undecidability of λK-conversion. In J. Bulloff, T. Holyoke, and S. Hahn, editors, *Foundations of Mathematics, Symposium Papers Commemorating the Sixtieth Birthday of Kurt Gödel*, pages 10–14. Springer-Verlag, Berlin, 1969.

[Curry, 1980] H. B. Curry. Some philosophical aspects of combinatory logic. In J. Barwise, H. J. Keisler, and K. Kunen, editors, *The Kleene Symposium*, pages 85–101. North-Holland Co., Amsterdam, 1980.

[Daalen, 1973] D. T. van Daalen. A description of Automath and some aspects of its language theory. In P. Braffort, editor, *Proceedings of the Symposium APLASM*, volume I, Orsay, 1973. Also in [Nederpelt *et al.*, 1994], Item A.3 pp. 101–126.

[Daalen, 1980] D. T. van Daalen. *The Language Theory of AUTOMATH*. PhD thesis, Technische Hogeschool Eindhoven, Netherlands, 1980. Main parts also in [Nederpelt *et al.*, 1994], Items A.6 (pp.163–200), B.6 (pp.303–312), C.5 (pp.493–654).

[Damas and Milner, 1982] L. Damas and R. Milner. Principal type-schemes for functional programming languages. In *Ninth Annual A.C.M. Symposium on the Principles of Programming Languages (POPL)*, pages 207–212. Association for Computing Machinery, New York, 1982.

[Damas, 1984] L. Damas. *Type Assignment in Programming Languages*. PhD thesis, Computer Science Dept., Univ. Edinburgh, U.K., 1984.

[Danos and Regnier, 1989] V. Danos and L. Regnier. The structure of multiplicatives. *Archive for Mathematical Logic*, 28:181–203, 1989.

[Davis, 1982] M. Davis. Why Gödel didn't have Church's Thesis. *Information and Control*, 54:3–24, 1982. Journal now called *Information and Computation*.

[Dekkers *et al.*, 1998] W. Dekkers, M. W. Bunder, and H. Barendregt. Completeness of the propositions-as-types interpretation of intuitionistic logic into illative combinatory logic. *Journal of Symbolic Logic*, 63:869–890, 1998.

[Dezani and Ermine, 1982] M. Dezani and F. Ermine. Maximal monoids of normal forms. *Fundamenta Informaticae*, 2:130–141, 1982.

[Dezani *et al.*, 1993] M. Dezani, S. Ronchi della Rocca, and M. Venturini Zilli, editors. *A Collection of Contributions in Honour of Corrado Böhm*, 1993. *Theoretical Computer Science* special issue Vol. 121.

[Dezani, 1975] M. Dezani. A type theory for λ-β-normal forms. Proceedings of the Symposium Informatica 75, Bled, Slovenia, 1975.

[Dezani, 1976] M. Dezani. Characterization of normal forms possessing inverse in the λ-β-η-calculus. *Theoretical Computer Science*, 2:323–337, 1976.

[Diller, 1968] J. Diller. Zur Berechenbarkeit primitiv-rekursiver Funktionale endlicher Typen. In H. A. Schmidt, editor, *Contributions to Mathematical Logic*, pages 109–120. North-Holland Co., Amsterdam, 1968. Proc. conf. Hannover 1966.

[Dragalin, 1968] A. Dragalin. The computation of primitive recursive terms of finite type, and primitive recusive realization. *Zapiski Naucnyh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta imeni V.A. Steklova, Akademii Nauk S.S.S.R. (L.O.M.I.)*, 8:32–45, 1968. In Russian.

[Drakengren, 1996] T. Drakengren. Uniqueness of Scott's reflexive domain in $\mathbf{P}\omega$. *Theoretical Computer Science*, 155:267–276, 1996.

[Droste and Göbel, 1993] M. Droste and R. Göbel. Universal domains and the amalgamation property. *Mathematical Structures in Computer Science*, 3:137–159, 1993.

[Dybyg, 1996] K. Dybyg. *The Scheme Programming Language*. Prentice-Hall, U.S.A., 1996.

[Egli and Constable, 1976] H. Egli and R. Constable. Computability concepts for programming language semantics. *Theoretical Computer Science*, 2:133–145, 1976.

[Eilenberg and Kelly, 1966] S. Eilenberg and G. M. Kelly. Closed categories. In *Proceedings of the Conference on Categorical Algebra, La Jolla, 1965*, pages 412–562. Springer-Verlag, Berlin, 1966.

[Enderton, 1995] H. B. Enderton. In memoriam: Alonzo Church 1903–1995. *Bulletin of Symbolic Logic*, 1:486–488, 1995.

[Enderton, 1998] H. B. Enderton. Alonzo Church and the Reviews. *Bulletin of Symbolic Logic*, 4:172–181, 1998.

[Engeler, 1981] E. Engeler. Algebras and combinators. *Algebra Universalis*, 13:389–392, 1981.

[Feferman, 1975a] S. Feferman. A language and axioms for explicit mathematics. In J. Crossley, editor, *Algebra and Logic*, volume 450 of *Lecture Notes in Mathematics*, pages 87–139. Springer-Verlag, Berlin, 1975.

[Feferman, 1975b] S. Feferman. Non-extensional type-free theories of partial operations and classifications, I. In J. Diller and G. Müller, editors, *Proof Theory Symposion, Kiel 1974*, volume 500 of *Lecture Notes in Mathematics*, pages 73–118. Springer-Verlag, Berlin, 1975.

[Feferman, 1977] S. Feferman. Categorical foundations and foundations of category theory. In R. Butts and J. Hintikka, editors, *Logic, Foundations of Mathematics and Computability Theory*, pages 149–169. D. Reidel Co., Dordrecht, Netherlands, 1977.

[Feferman, 1984] S. Feferman. Towards useful type-free theories, I. *Journal of Symbolic Logic*, 49:75–111, 1984. No Part II has been published.

[Felleisen *et al.*, 1987] M. Felleisen, D. Friedman, E. Kohlbecker, and B. Duba. A syntactic theory of sequential control. *Theoretical Computer Science*, 52:205–237, 1987.

[Felscher, 1986] W. Felscher. Dialogues as a foundation for intuitionistic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 3, pages 341–372. D. Reidel Co., Dordrecht, Netherlands, 1986.

[Feys, 1946a] R. Feys. La technique de la logique combinatoire. *Revue Philosophique de Louvain*, 44:74–103, 237–270, 1946.

[Feys, 1946b] R. Feys. Les méthodes récentes de déduction naturelle. *Revue Philosophique de Louvain*, 44:370–400, 1946.

[Fitch, 1936] F. B. Fitch. A system of formal logic without an analogue to the Curry W operator. *Journal of Symbolic Logic*, 1:92–100, 1936.

[Fitch, 1952] F. B. Fitch. *Symbolic Logic, an Introduction*. The Ronald Press, New York, 1952.

[Fitch, 1958] F. B. Fitch. Representation of sequential circuits in combinatory logic. *Philosophy of Science*, 25:263–279, 1958. Talk given 1957.

[Fitch, 1963] F. B. Fitch. The system $C\Delta$ of combinatory logic. *Journal of Symbolic Logic*, 28:87–97, 1963.

[Fitch, 1974] F. B. Fitch. *Elements of Combinatory Logic*. Yale Univ. Press, New Haven, U.S.A., 1974.

[Forster, 1995] T. L. Forster, editor. *Set Theory with a Universal Set*. Clarendon Press, Oxford, 1995. 1st edn. was 1992.

[Fourman and Grayson, 1982] M. Fourman and R. Grayson. Formal spaces. In A. S. Troelstra and D. van Dalen, editors, *The L. E. J. Brouwer Centenary Symposium*, pages 107–122. North-Holland Co., Amsterdam, 1982.

[Fox, 1966] L. Fox, editor. *Advances in Programming and Non-numerical Computation*. Pergamon Press, Oxford, 1966. Proceedings of a 1963 summer school.

[Frege, 1879] G. Frege. *Begriffsschrift*. Verlag Louis Nebert, Halle, 1879. Reprinted in *Begriffsschrift und Andere Aufsätze*, ed. by I. Angelelli, publ. 1964 by Georg Olms, Hildesheim, Germany. Engl. transl: *Begriffschrift*, in [Heijenoort, 1967], pp. 1–82.

[Frege, 1891] G. Frege. Funktion und Begriff. Vortrag Gehalten in der Sitzung vom 9. Januar 1891 der Jenaischen Gesellschaft für Medizin und Naturwissenschaft, 1891. English transl: *Function and Concept* in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach and M. Black, publ. Blackwell, Oxford, England, 1960 (2nd edn.), pp. 21–41. Same transl. also in *Gottlob Frege Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Blackwell 1984, pp. 137–156.

[Frege, 1893] G. Frege. *Grundgesetze der Arithmetik*. Verlag Hermann Pohle, Jena, 1893. Two vols. Reprinted as one vol. in 1962 by Georg Olms, Hildesheim, Germany, and in 1966 as No. 32 in series *Olms Paperbacks*. Partial English transl. in *The Basic Laws of Arithmetic, Exposition of the System*, ed. M. Furth, Univ. California Press, Berkeley, U.S.A., 1964.

[Freyd *et al.*, 1990] P. Freyd, P. Mulry, G. Rosolini, and D. Scott. Extensional PERs. In *Proceedings Fifth Annual IEEE Symposium on Logic In Computer Science*, pages 346–354. IEEE Computer Society, Los Alamitos, California, U.S.A., 1990.

[Freyd, 1990] P. J. Freyd. Recursive types reduced to inductive types. In *Proceedings Fifth Annual IEEE Symposium on Logic In Computer Science*, pages 498–507. IEEE Computer Society, Los Alamitos, California, U.S.A., 1990.

[Freyd, 1991] P. J. Freyd. Algebraically complete categories. In A. Carboni, M. Pedicchio, and G. Rosolini, editors, *Proceedings of the 1990 Como Category Theory Conference*, volume 1488 of *Lecture Notes in Mathematics*, pages 131–156. Springer-Verlag, 1991.

[Freyd, 1992] P. J. Freyd. Remarks on algebraically compact categories. In M. Fourman, P. Johnstone, and A. Pitts, editors, *Applications of Categories in Computer Science*, volume 177 of *London Math. Soc. Lecture Notes Series*, pages 95–106. Cambridge Univ. Press, 1992.

[Friedberg and Rogers, 1959] R. M. Friedberg and H. Rogers. Reducibilities and completeness for sets of integers. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 5:117–125, 1959.

[Friedman, 1975] H. Friedman. Equality between functionals. In R. Parikh, editor, *Logic Colloquium, Symposium on Logic held at Boston, 1972–73*, volume 453 of *Lecture Notes in Mathematics*, pages 22–37. Springer-Verlag, Berlin, 1975.

[Friedman, 1978] H. Friedman. Classically and intuitionistically provably recursive functions. In G. Müller and D. Scott, editors, *Higher Set Theory*, volume 669 of *Lecture Notes in Mathematics*, pages 21–27. Springer-Verlag, 1978.

[Gallier, 1990] J. H. Gallier. On Girard's "Candidats de reductibilité". In P. Odifreddi, editor, *Logic and Computer Science*, volume 31 of *APIC Studies in Data Processing*, pages 123–203. Academic Press, London, 1990.

[Gandy, 1977] R. O. Gandy. The simple theory of types. In R. Gandy and M. Hyland, editors, *Logic Colloquium '76*, pages 173–181. North-Holland Co., Amsterdam, 1977.

[Gandy, 1988] R. O. Gandy. The confluence of ideas in 1936. In R. Herken, editor, *The Universal Turing Machine: A Half-Century Survey*, pages 51–102. Oxford Univ. Press, England, 1988. (2nd edn. 1995, publ. Springer-Verlag).

[Gandy, 1993] R. O. Gandy. Dialogues, Blass games, sequentiality for objects of finite type. Unpublished manuscript, 1993.

[Gentzen, 1935] G. Gentzen. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1935. English transl: *Investigations into logical deduction*, in *The Collected Papers of Gerhard Gentzen*, ed. by M. Szabo, North-Holland Co., Amsterdam, 1969.

[Giannini and Longo, 1984] P. Giannini and G. Longo. Effectively given domains and lambda calculus semantics. *Information and Control*, 62:36–63, 1984. Journal now *Information and Computation*.

[Gierz *et al.*, 1980] G. Gierz, K. Hofmann, K. Keimel, J. Lawson, M. Mislove, and D. Scott. *A Compendium of Continuous Lattices*. Springer-Verlag, Berlin, 1980.

[Gilmore, 1963] P. C. Gilmore. An abstract computer with a LISP-like machine language without a label operator. In P. Braffort and D. Hirschberg, editors, *Computer Programming and Formal Systems*, pages 71–86. North-Holland Co., Amsterdam, 1963.

[Girard *et al.*, 1989] J.-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types*. Cambridge Univ. Press, England, 1989.

[Girard, 1971] J.-Y. Girard. Une extension de l'interprétation de Gödel à l'analyse, et son application à l'élimination des coupures dans l'analyse et la théorie des types. In J. E. Fenstad, editor, *Proceedings of the Second Scandinavian Logic Symposium*, pages 63–92. North-Holland Co., Amsterdam, 1971.

[Girard, 1972] J.-Y. Girard. Interprétation fonctionnelle et élimination des coupures de l'arithmétique d'ordre supérieur. Thèse de Doctorat d'État, Univ. Paris VII, Place Jussieu, Paris, 1972.

[Girard, 1981] J.-Y. Girard. Π^1_2-logic, part I: dilators. *Annals of Mathematical Logic*, 21:75–219, 1981.

[Girard, 1986] J.-Y. Girard. The system F of variable types, fifteen years later. *Theoretical Computer Science*, 45:159–192, 1986.

[Girard, 1987a] J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50:1–102, 1987.

[Girard, 1987b] J.-Y. Girard. Multiplicatives. *Rendiconti del Seminario Matematico dell' Università e del Politecnico di Torino*, Fascicolo speciale: "Logic and Computer Science: New Trends and Applications":11–34, 1987. Edited by G. Lolli.

[Girard, 1988] J.-Y. Girard. Normal functors, power series and λ-calculus. *Annals of Mathematical Logic*, 37:129–177, 1988.

[Girard, 1989a] J.-Y. Girard. Geometry of interaction I: interpretation of system F. In R. Ferro, C. Bonotto, S. Valentini, and A. Zanardo, editors, *Logic Colloquium '88*, pages 221–260. North-Holland Co., Amsterdam, 1989.

[Girard, 1989b] J.-Y. Girard. Towards a geometry of interaction. In J. W. Gray and A. Scedrov, editors, *Categories in Computer Science and Logic*, volume 92 of *Contemporary Mathematics*, pages 69–108. American Mathematical Society, Providence, Rhode Island, 1989.

[Girard, 1998] J.-Y. Girard. On the meaning of logical rules I: syntax vs. semantics. In U. Berger and H. Schwichtenberg, editors, *Computational Logic*, pages 215–272. Springer-Verlag, Berlin, 1998. Proc. of meeting 1997.

[Girard, 2000] J.-Y. Girard. On the meaning of logical rules II: multiplicative/additive case. In F. L. Bauer and R. Steinbrüggen, editors, *Foundations of Secure Computation*. IOS Press, Amsterdam, 2000. Preprint 1998, Inst. de Math. de Luminy, Marseilles.

[Girard, 2001] J.-Y. Girard. Locus solum. *Mathematical Structures in Computer Science*, 11:299–506, 2001.

[Givant and Andréka, 2002] S. Givant and H. Andréka. Groups and algebras of binary relations. *Bulletin of Symbolic Logic*, 8:38–64, 2002.

[Gödel, 1958] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287, 1958. Engl. transl: *On a hitherto unexploited extension of the finitary standpoint*, in *Journal of Philosophical Logic* 9 (1980) pp. 133–142. Two other Engl. transls. with notes: in [Gödel, 1990, pp. 217–251, 271–280].

[Gödel, 1990] K. Gödel. *Collected Works, Vol. 2: Publications 1938–1974*. Oxford Univ. Press, New York and Oxford, 1990. (Eds. S. Feferman et al.).

[Gödel, 1995] K. Gödel. *Collected Works, Vol. 3: Unpublished Essays and Lectures*. Oxford Univ. Press, New York and Oxford, 1995. (Eds. S. Feferman et al.).

[Gonthier *et al.*, 1992] G. Gonthier, M. Abadi, and J.-J. Levy. The geometry of optimal lambda reduction. In *Nineteenth Annual A.C.M. Symposium on the Principles of Programming Languages (POPL)*, pages 15–26, Albuquerque, New Mexico, 1992. Association for Computing Machinery, New York.

[Goodman, 1970] N. D. Goodman. A theory of constructions equivalent to arithmetic. In A. Kino, J. Myhill, and R. Vesley, editors, *Intuitionism and Proof Theory*, pages 101–120. North-Holland Co., Amsterdam, 1970.

[Gordon *et al.*, 1979] M. Gordon, R. Milner, and C. Wadsworth. *Edinburgh LCF. A Mechanical Logic of Computation*, volume 78 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1979.

[Gordon, 2000] M. Gordon. From LCF to HOL: a short history. In G. Plotkin, C. Stirling, and M. Tofte, editors, *Proof, Language and Interaction: Essays in Honour of Robin Milner*. M.I.T. Press, Cambridge, Mass., U.S.A., 2000.

[Grattan-Guinness, 2000] I. Grattan-Guinness. *The Search for Mathematical Roots, 1870–1940*. Princeton Univ. Press, U.S.A., 2000.

[Griffin, 1990] T. Griffin. The formulae-as-types notion of control. In *Seventeenth Annual A.C.M. Symposium on the Principles of Programming Languages (POPL)*, pages 47–58. Association for Computing Machinery, New York, 1990.

[Gunter and Jung, 1990] C. Gunter and A. Jung. Coherence and consistency in domain. *Journal of Pure and Applied Algebra*, 63:49–66, 1990.

[Gunter and Mitchell, 1994] C. Gunter and J. C. Mitchell. *Theoretical Aspects of Object-Oriented Programming: Types, Semantics and Language Design*. M.I.T. Press, Cambridge, Mass., U.S.A., 1994.

[Gunter, 1987] C. Gunter. Universal profinite domains. *Information and Computation*, 72:1–30, 1987.

[Halmos, 1962] P. Halmos. *Algebraic Logic*. Chelsea Publishing Co., New York, 1962.

[Hanatani, 1966] Y. Hanatani. Calculabilité des fonctionnels recursives primitives de type fini sur les nombres naturels. *Annals of the Japan Association for the Philosophy of Science*, 3:19–30, 1966. (Revised version: *Calculability of the primitive recursive functionals of finite type over the natural numbers*, in *Proof Theory Symposion, Kiel 1974*, ed. by J. Diller and G. Müller, Lecture Notes in Mathematics 500, Springer-Verlag, Berlin, 1975, pp. 152–163).

[Hancock and Martin-Löf, 1975] P. Hancock and P. Martin-Löf. Syntax and semantics of the language of primitive recursive functions. Technical Report 3, Univ. Stockholm, 1975.

[Hankin, 1994] C. Hankin. *Lambda Calculi*. Clarendon Press, Oxford, England, 1994.

[Harper *et al.*, 1993] R. Harper, F. Honsell, and G. D. Plotkin. A framework for defining logics. *Journal of the ACM*, 40:143–184, 1993.

[Heijenoort, 1967] J. van Heijenoort. *From Frege to Gödel: A Source Book in Mathematical Logic 1879–1931*. Harvard University Press, Cambridge, Mass., U.S.A., 1967.

[Henkin, 1950] L. Henkin. Completeness in the theory of types. *Journal of Symbolic Logic*, 15:81–91, 1950.

[Heyting, 1956] A. Heyting. *Intuitionism, an Introduction*. North-Holland Co., Amsterdam, 1956. (3rd edn. 1973).

[Heyting, 1958] A. Heyting. Intuitionism in mathematics. In R. Klibansky, editor, *Philosophy in the Mid-century*, pages 101–115. La Nuova Italia, Firenze, 1958.

[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der Theoretischen Logik*. Springer-Verlag, Berlin, 1928. (English transl. of 2nd (1938) edn., with some corrections: *Principles of Mathematical Logic*, Chelsea Publ. Co., New York 1950).

[Hinata, 1967] S. Hinata. Calculability of primitive recursive functionals of finite type. *Science Reports of the Tokyo Kyoiku Daigaku, Section A*, 9(226):42–59, 1967.

[Hindley and Longo, 1980] J. R. Hindley and G. Longo. Lambda-calculus models and extensionality. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 26:289–310, 1980.

[Hindley and Seldin, 1980] J. R. Hindley and J. P. Seldin, editors. *To H. B. Curry, Essays on Combinatory Logic, Lambda Calculus and Formalism*. Academic Press, London, 1980.

[Hindley and Seldin, 1986] J. R. Hindley and J. P. Seldin. *Introduction to Combinators and λ-calculus*. Cambridge Univ. Press, England, 1986.

[Hindley *et al.*, 1972] J. R. Hindley, B. Lercher, and J. P. Seldin. *Introduction to Combinatory Logic*. Cambridge Univ. Press, England, 1972.
[Hindley, 1969a] J. R. Hindley. An abstract form of the Church-Rosser theorem, I. *Journal of Symbolic Logic*, 34:545–560, 1969. From author's PhD thesis, Univ. Newcastle upon Tyne, England, 1964.
[Hindley, 1969b] J. R. Hindley. The principal type-scheme of an object in combinatory logic. *Transactions of the American Mathematical Society*, 146:29–60, 1969.
[Hindley, 1978a] J. R. Hindley. Reductions of residuals are finite. *Transactions of the American Mathematical Society*, 240:345–361, 1978.
[Hindley, 1978b] J. R. Hindley. Standard and normal reductions. *Transactions of the American Mathematical Society*, 241:253–271, 1978.
[Hindley, 1982] J. R. Hindley. The simple semantics for Coppo-Dezani-Sallé types. In M. Dezani and U. Montanari, editors, *International Symposium on Programming, Proceedings 1982*, volume 137, pages 212–226. Springer-Verlag, Berlin, 1982.
[Hindley, 1997] J. R. Hindley. *Basic Simple Type Theory*. Cambridge Univ. Press, England, 1997.
[Hindley, 2008] J. R. Hindley. M. h. newman's typability algorithm for lambda-calculus. *Journal of Logic and Computation*, 2008.
[Hodges, 1983] A. Hodges. *Alan Turing: The Enigma*. Burnett Books, London, 1983. Also published 1988 by Vintage, London.
[Holmes, 1991] M. R. Holmes. Systems of combinatory logic related to Quine's 'New Foundations'. *Annals of Pure and Applied Logic*, 53:103–133, 1991.
[Holmes, 1995] M. R. Holmes. Untyped λ-calculus with relative typing. In M. Dezani and G. Plotkin, editors, *Typed Lambda Calculi and Applications, Second International Conference, TLCA '95*, number 902 in Lecture Notes in Computer Science, pages 235–248. Springer-Verlag, 1995.
[Howard, 1970] W. A. Howard. Assignment of ordinals to terms for primitive recursive functionals of finite type. In A. Kino, J. Myhill, and R. Vesley, editors, *Intuitionism and Proof Theory*, pages 443–458. North-Holland Co., Amsterdam, 1970. Proc. conf. at Buffalo, N.Y., 1968.
[Howard, 1980] W. A. Howard. The formulæ-as-types notion of construction. In Hindley and Seldin [1980], pages 479–490. Manuscript circulated 1969.
[Huet, 1980] G. Huet. Confluent reductions. *Journal of the Association for Computing Machinery*, 27:797–821, 1980.
[Huet, 1990a] G. Huet. Cartesian closed categories and lambda-calculus. In Huet [1990b], pages 7–23.
[Huet, 1990b] G. Huet, editor. *Logical Foundations of Functional Programming*. Univ. Texas at Austin Year of Programming. Addison-Wesley, Reading, Mass., U.S.A., 1990.
[Huet, 1993] G. Huet. An analysis of Böhm's theorem. *Theoretical Computer Science*, 121 (Böhm Volume):154–167, 1993.
[Husserl, 1922] E. Husserl. *Logische Untersuchungen*. Max Niemeyer, Halle, Germany, 3rd edition, 1922. (1st edn. was 1900–1901).
[Hyland and Ong, 2000] J. M. E. Hyland and C.-H. L. Ong. On full abstraction for PCF: I. Models, observables and the full abstraction problem, II. Dialogue games and innocent strategies, III. A fully abstract and universal game model. *Information and Computation*, 163:285–408, 2000.
[Hyland, 1975] J. M. E. Hyland. A survey of some useful partial order relations on terms of the lambda calculus. In Böhm [1975], pages 83–95.
[Hyland, 1976] J. M. E. Hyland. A syntactic characterization of the equality in some models for the lambda calculus. *Journal of the London Mathematical Society, Series 2*, 12:361–370, 1976.
[Hyland, 1982] J. M. E. Hyland. The effective topos. In A. S. Troelstra and D. van Dalen, editors, *The L.E.J. Brouwer Centenary Symposium*, pages 165–216. North-Holland Co., Amsterdam, 1982.
[Hyland, 1988] J. M. E. Hyland. A small complete category. *Annals of Pure and Applied Logic*, 40:135–165, 1988.
[Hyland, 1991] J. M. E. Hyland. First steps in synthetic domain theory. In A. Carboni, editor, *Conference on Category Theory*, number 1488 in Lecture Notes in Mathematics, pages 131–156. Springer-Verlag, 1991.

[Jacobs, 1989] B. Jacobs. The inconsistency of higher order extensions of Martin-Löf's type theory. *Journal of Philosophical Logic*, 18(4):399–422, 1989.

[Johnstone *et al.*, 1978] P. Johnstone, R. Paré, R. Rosebrugh, S. Schumacher, R. Wood, and G. Wraith. *Indexed Categories and their Applications*. Number 661 in Lecture Notes in Mathematics. Springer-Verlag, 1978.

[Johnstone, 1982] P. Johnstone. *Stone Spaces*. Number 3 in Cambridge Studies in Advanced Mathematics. Cambridge Univ. Press, 1982.

[Joyal, 1977] A. Joyal. Remarques sur la théorie des jeux a deux personnes. *Gazette des Sciences Mathématiques du Quebec*, 1(4), 1977.

[Jung (editor), 1996] A. Jung (editor). Domains and denotational semantics: history, accomplishments and open problems. *Bulletin of the European Association for Theoretical Computer Science*, 59:227–256, 1996.

[Kahn and Plotkin, 1978] G. Kahn and G. D. Plotkin. Domaines concrèts. Rapport 336, IRIA Laboria, 1978. English transl. in *Theoretical Computer Science* 121 (Böhm Volume 1993), pp. 187–277.

[Kahn, 1988] G. Kahn. Natural semantics. In K. Fuchi and M. Nivat, editors, *Programming of Future Generation Computers*, pages 237–258. Elsevier, Amsterdam, 1988.

[Kalman, 1983] J. Kalman. Condensed detachment as a rule of inference. *Studia Logica*, 42:443–451, 1983.

[Kamareddine and Nederpelt, 1993] F. Kamareddine and R. P. Nederpelt. On stepwise explicit substitutions. *International Journal of Foundations of Computer Science*, 4:197–240, 1993.

[Kamareddine and Rios, 1997] F. Kamareddine and A. Rios. Extending a λ-calculus with explicit substitution which preserves strong normalisation into a confluent calculus on open terms. *Journal of Functional Programming*, 7:395–420, 1997.

[Kamareddine and Rios, 2000] F. Kamareddine and A. Rios. Relating the lambda-sigma and lambda-s styles of explicit substitutions. *Journal of Logic and Computation*, 10:349–380, 2000.

[Kamareddine *et al.*, 2002] F. Kamareddine, T. Laan, and R. P. Nederpelt. Types in logic and mathematics before 1940. *Bulletin of Symbolic Logic*, 8:185–245, 2002.

[Kanda, 1979] A. Kanda. Fully effective solutions of recursive domain equations. In J. Beçvar, editor, *Mathematical Foundations of Computer Science*, volume 74 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1979.

[Kathail, 1990] V. Kathail. *Optimal interpreters for lambda-calculus based functional languages*. PhD thesis, Massachusetts Institute of Technology, 1990.

[Kearns, 1973] J. T. Kearns. The completeness of combinatory logic with discriminators. *Notre Dame Journal of Formal Logic*, 14:323–330, 1973.

[Kelly and Mac Lane, 1972] G. M. Kelly and S. Mac Lane. Coherence in closed categories. *Journal of Pure and Applied Algebra*, 1:97–140, 1972.

[Kleene and Rosser, 1935] S. C. Kleene and J. B. Rosser. The inconsistency of certain formal logics. *Annals of Mathematics, Series 2*, 36:630–636, 1935.

[Kleene, 1936] S. C. Kleene. λ-definability and recursiveness. *Duke Mathematical Journal*, 2:340–353, 1936.

[Kleene, 1945] S. C. Kleene. On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic*, 10:109–124, 1945.

[Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics*. Van Nostrand, New York, 1952.

[Kleene, 1978] S. C. Kleene. Recursive functionals and quantifiers of finite types revisited I. In J. E. Fenstad, R. O. Gandy, and G. E. Sacks, editors, *Generalized Recursion Theory II*, pages 185–222. North-Holland, Amsterdam, 1978.

[Kleene, 1980] S. C. Kleene. Recursive functionals and quantifiers of finite types revisited II. In J. Barwise, H. J. Keisler, and K. Kunen, editors, *The Kleene Symposium*, pages 1–29. North-Holland, Amsterdam, 1980.

[Kleene, 1981] S. C. Kleene. Origins of recursive function theory. *Annals of the History of Computing*, 3:52–67, 1981. Preliminary version in: *20th IEEE Symposium on the Foundations of Computer Science*, publ. by IEEE Computer Society, Los Alamitos, California, U.S.A., 1979, pp. 371–382.

[Kleene, 1982] S. C. Kleene. Recursive functionals and quantifiers of finite types revisited III. In G. Metakides, editor, *Patras Logic Symposium*, pages 1–40. North-Holland Co., Amsterdam, 1982.

[Kleene, 1985] S. C. Kleene. Unimonotone functions of finite types (Recursive functionals and quantifiers of finite types revisited IV). In A. Nerode and R. A. Shore, editors, *Recursion Theory*, pages 119–138. American Mathematical Society, Providence, Rhode Island, 1985.

[Klement, 2003] K. C. Klement. Russell's 1903–1905 anticipation of the lambda calculus. *History and Philosophy of Logic*, 24:15–37, 2003.

[Kline, 1951] G. L. Kline. Review of: Yanovskaya, S. A.: Osnovaniya matematiki i matematicheskaya logika (Foundations of mathematics and mathematical logic). *Journal of Symbolic Logic*, 16:46–48, 1951. Article reviewed: pp. 9–50 of *Matematika v SSSR za Tridkat let 1917–1947* (Mathematics in the USSR in the 30 Years 1917–1947), in Russian, publ. OGIZ, Moscow, 1948.

[Klop *et al.*, 1993] J. W. Klop, V. van Oostrom, and F. van Raamsdonk. Combinatory reduction systems: introduction and survey. *Theoretical Computer Science*, 121(1–2):279–308, 1993.

[Klop, 1980] J. W. Klop. *Combinatory Reduction Systems*. PhD thesis, Univ. Utrecht, 1980. Publ. by Mathematisch Centrum, 413 Kruislaan, Amsterdam.

[Klop, 1992] J. W. Klop. Term Rewriting Systems. In S. Abramsky, D. Gabbay, and T. Maibaum, editors, *Handbook of Logic in Computer Science, Volume 2: Background, Computational Structures*, pages 1–116. Oxford Univ. Press, England, 1992.

[Komori, 1990] Y. Komori. λ-calculus with simultaneous substitution as a primitive symbol. In *Proceedings of the 13th Symposium on Semigroups*, pages 20–26. Kyoto Sangyo Univ., 1990. (Symposium in 1989).

[Koymans, 1982] C. P. J. Koymans. Models of the lambda calculus. *Information and Control*, 52:306–332, 1982. (Journal now called *Information and Computation*).

[Kreisel, 1958] G. Kreisel. Constructive mathematics, 1958. Mimeographed notes of a course given at Stanford Univ., 1958–59.

[Kreisel, 1959] G. Kreisel. Interpretation of analysis by means of constructive functionals of finite type. In A. Heyting, editor, *Constructivity in Mathematics*, pages 101–128. North-Holland Co., Amsterdam, 1959.

[Kreisel, 1962] G. Kreisel. Foundations of intuitionistic logic. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology, and the Philosophy of Science*, pages 198–210, Stanford, 1962.

[Kreisel, 1965] G. Kreisel. Mathematical logic. In T. L. Saaty, editor, *Lectures on Modern Mathematics*, volume 3, pages 95–195. Wiley, New York, 1965.

[Krivine, 1990] J.-L. Krivine. *Lambda-calcul, Types et Modèles*. Masson, Paris, 1990. English transl: *Lambda-Calculus, Types and Models*, Ellis-Horwood (U.S.A.) and Prentice-Hall (U.K.), 1993.

[Kuzichev, 1980] A. S. Kuzichev. Sequential systems of λ-conversion and of combinatory logic. In Hindley and Seldin [1980], pages 141–155.

[Kuzichev, 1983] A. S. Kuzichev. Set theory in type-free combinatorially complete systems. *Vestnik, Moscow State University, Series Math./Mechan.*, pages 36–42, 1983. In Russian.

[Kuzichev, 1999] A. S. Kuzichev. A version of formalization of Cantor's set theory. *Doklady Mathematics*, 60:424–426, 1999.

[Laan, 1997] T. D. L. Laan. *The Evolution of Type Theory in Logic and Mathematics*. PhD thesis, Tech. Univ. Eindhoven, Netherlands, 1997.

[Lafont and Streicher, 1991] Y. Lafont and T. Streicher. Games semantics for linear logic. In *Proceedings Sixth Annual IEEE Symposium on Logic In Computer Science*, pages 43–50. IEEE Computer Society, Los Alamitos, California, U.S.A., 1991.

[Lamarche, 1992] F. Lamarche. Sequentiality, games and linear logic, 1992. Unpublished paper, from a lecture to the CLiCS Symposium (Aarhus Univ., March 23–27, 1992).

[Lambek and Findlay, 1955] J. Lambek and G. D. Findlay. Calculus of bimodules, 1955. Unpublished manuscript.

[Lambek, 1958] J. Lambek. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170, 1958.

[Lambek, 1961] J. Lambek. On the calculus of syntactic types. In R. Jakobson, editor, *Structure of Language and its Mathematical Aspects*, number 12 in Proceedings of Symposia in Applied Mathematics, pages 56–68. American Math. Soc. Providence, R.I., U.S.A., 1961.

[Lambek, 1968] J. Lambek. Deductive systems and categories, I: syntactic calculus and residuated categories. *Mathematical Systems Theory*, 2(4):287–318, 1968.

[Lambek, 1969] J. Lambek. Deductive systems and categories, II: standard constructions and closed categories. In P. Hilton, editor, *Category Theory, Homology Theory and their Applications II*, volume 86 of *Lecture Notes in Mathematics*, pages 76–122. Springer-Verlag, Berlin, 1969.

[Lambek, 1972] J. Lambek. Deductive systems and categories, III: cartesian closed categories, intuitionist propositional calculus, and combinatory logic. In W. F. Lawvere, editor, *Toposes, Algebraic Geometry and Logic*, volume 274 of *Lecture Notes in Mathematics*, pages 57–82. Springer-Verlag, Berlin, 1972.

[Lambek, 1974] J. Lambek. Functional completeness of cartesian categories. *Annals of Mathematical Logic*, 6:259–292, 1974.

[Lambek, 1980a] J. Lambek. From λ-calculus to cartesian closed categories. In Hindley and Seldin [1980], pages 375–402.

[Lambek, 1980b] J. Lambek. From types to sets. *Advances in Mathematics*, 36:113–164, 1980.

[Lambek, 1988] J. Lambek. Categorial grammars and natural language structures. In R. Oehrle, E. Bach, and D. Wheeler, editors, *Categorial and Categorical Grammars*. D. Reidel Co., Dordrecht, Netherlands, 1988.

[Lambek, 1995] J. Lambek. Bilinear logic in algebra and linguistics. In J.-Y. Girard, Y. Lafont, and L. Regnier, editors, *Advances in Linear Logic*, pages 43–59. Cambridge Univ. Press, England, 1995. (Proc. 1993 Workshop on Linear Logic, Ithaca, N.Y., U.S.A.).

[Lamping, 1990] J. Lamping. An algorithm for optimal lambda calculus reduction. In *Seventeenth Annual A.C.M. Symposium on the Principles of Programming Languages (POPL)*, pages 16–30. Association for Computing Machinery, New York, 1990.

[Landin, 1964] P. J. Landin. The mechanical evaluation of expressions. *The Computer Journal*, 6:308–320, 1964. Actually appeared 1963.

[Landin, 1965] P. J. Landin. A correspondence between ALGOL 60 and Church's lambda notation. *Communications of the ACM*, 8:89–101, 158–165, 1965.

[Landin, 1966a] P. J. Landin. A lambda-calculus approach. In L. Fox, editor, *Advances in Programming and Non-numerical Computation*, pages 97–141. Pergamon Press, England, 1966.

[Landin, 1966b] P. J. Landin. The next 700 programming languages. *Communications of the ACM*, 9(3):157–166, 1966.

[Läuchli, 1965] H. Läuchli. Intuitionistic propositional calculus and definably non-empty terms. *Journal of Symbolic Logic*, 30:263, 1965. Abstract only.

[Läuchli, 1970] H. Läuchli. An abstract notion of realizability for which intuitionistic predicate calculus is complete. In A. Kino, J. Myhill, and R. Vesley, editors, *Intuitionism and Proof Theory*, pages 227–234. North-Holland Co., Amsterdam, 1970. Proc. 1968 conference at Buffalo, N.Y.

[Lawvere, 1963] F. W. Lawvere. *Functorial semantics of algebraic theories*. PhD thesis, Columbia Univ., 1963.

[Lawvere, 1964] F. W. Lawvere. An elementary theory of the category of sets. *Proceedings of the National Academy of Sciences of the U.S.A.*, 52(6):1506–1511, 1964. Extended abstract of a typescript with the same title (1964, Univ. Chicago).

[Lawvere, 1966] F. W. Lawvere. The category of categories as a foundation for Mathematics. In *Proceedings of the La Jolla conference on Categorical Algebra*, pages 1–20. Springer-Verlag, 1966.

[Lawvere, 1969a] F. W. Lawvere. Adjointness in Foundations. *Dialectica*, 23:281–296, 1969.

[Lawvere, 1969b] F. W. Lawvere. Diagonal arguments and cartesian closed categories. In P. Hilton, editor, *Category Theory, Homology Theory and their Applications II*, volume 92 of *Lecture Notes in Mathematics*, pages 134–145. Springer-Verlag, Berlin, 1969.

[Lawvere, 1970a] F. W. Lawvere. Equality in hyperdoctrines and comprehension schema as an adjoint functor. In A. Heller, editor, *Applications of Categorical Algebra*, volume 17 of *Proceedings of Symposia on Pure Mathematics*, pages 1–14. American Math. Soc., Providence, R.I., U.S.A., 1970. Proc. of a symposium in New York 1968.

[Lawvere, 1970b] F. W. Lawvere. Quantifiers and sheaves. In *Actes du Congrès International des Mathèmatiques, Nice 1970*, pages 329–334, 1970. tome I.

[Lawvere, 1972] F. W. Lawvere. Category theory over a base topos, 1972. Unpubl. lecture notes.

[Lawvere, 1976] F. W. Lawvere. Variable quantities and variable structures in topoi. In *Algebra, Topology and Category Theory. A collection of Papers in Honor of Samuel Eilenberg*, pages 101–131. Academic Press, London and New York, 1976.

[Lawvere, 1979] F. W. Lawvere. Categorical dynamics. In *Proceedings of the Aarhus Open House on Topos Theoretic Methods in Geometry, May 1978*, Aarhus, Denmark, 1979.

[Lawvere, 1996] F. W. Lawvere. Adjoints in and among bicategories. In A. Ursini and P. Aglianò, editors, *Logic and Algebra. Proceedings of the 1994 Siena Conference in memory of Roberto Magari*, volume 180 of *Lecture Notes in Pure and Applied Algebra*, pages 181–189. Marcel Dekker, Inc., New York, Basel and Hong Kong, 1996.

[Leivant, 1983] D. Leivant. Polymorphic type inference. In *Tenth Annual A.C.M. Symposium on the Principles of Programming Languages (POPL)*, pages 88–98. Association for Computing Machinery, New York, 1983. Cf. also pp. 155–166.

[Leivant, 1990] D. Leivant. Contracting proofs to programs. In P. Odifreddi, editor, *Logic and Computer Science*, volume 31 of *APIC Studies in Data Processing*, pages 279–327. Academic Press, London, 1990.

[Lemmon *et al.*, 1957] E. J. Lemmon, C. A. Meredith, D. Meredith, A. N. Prior, and I. Thomas. Calculi of pure strict implication. Informally distributed, 1957. Publ. 1969 in *Philosophical logic*, ed. by Davis and Hockney and Wilson, D. Reidel Co., Dordrecht, Netherlands, pp. 215–250.

[Lercher, 1966] B. Lercher. Review of: K. Menger: the algebra of functions: past, present, future. *Journal of Symbolic Logic*, 31:272, 1966.

[Leśniewski, 1929] S. Leśniewski. Grundzüge eines neuen Systems der Grundlagen der Mathematik. *Fundamenta Mathematicæ*, 14:1–81, 1929.

[Lévy, 1976] J.-J. Lévy. An algebraic interpretation of $\lambda\beta$K-calculus, and an application of a labelled λ-calculus. *Theoretical Computer Science*, 2:97–114, 1976. Preliminary version 1975 in [Böhm, 1975] pp. 147–165.

[Lévy, 1978] J.-J. Lévy. *Réductions correctes et optimales dans le λ-calcul.* Thèse de Doctorat d'Etat, Univ. Paris VII, Paris, 1978.

[Loader, 2001] R. Loader. Finitary PCF is not decidable. *Theoretical Computer Science*, 266:341–364, 2001.

[Longley, 2001] J. R. Longley. Notions of computability at higher types, I. In *Logic Colloquium 2000*, volume 19 of *Lecture Notes in Logic*, pages 32–142. A. K. Peters, 2001.

[Longo and Moggi, 1992] G. Longo and E. Moggi. Constructive natural deduction and its modest interpretation. *Mathematical Structures in Computer Science*, 1:215–254, 1992.

[Longo *et al.*, 1993] G. Longo, K. Milsted, and S. Soloviev. The genericity theorem and parametricity in the polymorphic λ-calculus. *Theoretical Computer Science*, 121:323–349, 1993.

[Longo, 1987] G. Longo. From numbered sets to type theories. *Rendiconti del Seminario Matematico dell'Università e del Politecnico di Torino*, Fascicolo speciale: "Logic and Computer Science: New Trends and Applications":41–73, 1987. Ed. by G. Lolli.

[Longo, 1995] G. Longo. Parametric and type-dependent polymorphism. *Fundamenta Informaticae*, 22:69–92, 1995.

[Lorenz, 2001] K. Lorenz. Basic objectives of dialogue logic in historical perspective. *Synthese*, 127:255–263, 2001.

[Lorenzen, 1955] P. Lorenzen. *Einführung in die Operative Logik und Mathematik.* Springer-Verlag, Berlin, Göttingen, Heidelberg, 1955.

[Lorenzen, 1961] P. Lorenzen. Ein dialogisches Konstructivitätskriterium. In *Infinitistic Methods*, pages 193–200. Pergamon Press and PWN, Oxford and Warsaw, 1961.

[Löwenheim, 1940] L. Löwenheim. Einkleidung der Mathematik in Schröderschen Relativkalkul. *Journal of Symbolic Logic*, 5:1–15, 1940.

[Łukasiewicz and Tarski, 1930] J. Łukasiewicz and A. Tarski. Untersuchungen über den Aussagenkalkül. *Comptes Rendus des Séances de la Societé des Sciences et des Lettres de Varsovie*, 23:30–50, 1930. Engl. transl: *Investigations into the sentential calculus*, in *Logic, Semantics, Metamathematics*, by A. Tarski, Clarendon Press, Oxford, 1956, pp. 38–59.

[Łukasiewicz, 1939] J. Łukasiewicz. Der äquivalenzenkalkül. *Collectanea Logica*, 1:145–169, 1939. Journal vol. never appeared. Engl. transl: *The equivalential calculus*, in *Polish Logic*, ed. by S. McCall, Clarendon Press, Oxford, pp. 88–115, and in *Jan Łukasiewicz Selected Works*, ed. by L. Borkowski, North-Holland Co., Amsterdam, 1970, pp. 250–277.

[MacQueen *et al.*, 1984] D. MacQueen, G. D. Plotkin, and R. Sethi. An ideal model for recursive polymorphic types. In *Eleventh Annual A.C.M. Symposium on the Principles of Programming Languages (POPL)*, pages 165–174. Association for Computing Machinery, New York, 1984.

[Mann, 1975] C. R. Mann. The connection between equivalence of proofs and cartesian closed categories. *Proceedings of the London Mathematical Society*, 31(3):289–310, 1975.

[Manzano, 1997] M. Manzano. Alonzo Church: his life, his work and some of his miracles. *History and Philosophy of Logic*, 18:211–232, 1997.

[Martin-Löf, 1971a] P. Martin-Löf. On the strength of intuitionistic reasoning, 1971. Unpublished contribution to Symposium on Perspectives in the Philosophy of Mathematics at 4th International Congress for Logic, Methodology and Philosophy of Science, Bucharest.

[Martin-Löf, 1971b] P. Martin-Löf. A Theory of Types, 1971. Informally circulated.

[Martin-Löf, 1972a] P. Martin-Löf. Infinite terms and a system of natural deduction. *Compositio Mathematica*, 24(1):93–103, 1972.

[Martin-Löf, 1972b] P. Martin-Löf. An Intuitionistic Theory of Types. Technical report, Dept. of Mathematics, Univ. Stockholm, 1972. Published in *Twenty-Five Years of Constructive Type Theory*, edited by Giovanni Sambin and Jan Smith, Clarendon Press, Oxford, 1998, pp. 127–172.

[Martin-Löf, 1975] P. Martin-Löf. An intuitionistic theory of types: Predicative part. In H. Rose and J. Shepherdson, editors, *Logic Colloquium '73*, pages 73–118. North-Holland Co., Amsterdam, 1975.

[Martin-Löf, 1982] P. Martin-Löf. Constructive mathematics and computer programming. In L. J. Cohen, J. Łos, H. Pfeiffer, and K-P. Podewski, editors, *Logic, Methodology and Philosophy of Science, VI*, pages 153–175. North-Holland Co., Amsterdam, 1982.

[Martin-Löf, 1983] P. Martin-Löf. The domain interpretation of type theory. In K. Karlsson and K. Petersson, editors, *Workshop on the Semantics of Programming Languages, Abstracts and Notes*. Programming Methodology Group, Chalmers Univ. Technology and Univ. Göteborg, August 1983.

[Martin-Löf, 1984] P. Martin-Löf. *Intuitionistic Type Theory*. Studies in Proof Theory. Bibliopolis, Napoli, Italy, 1984. Notes by Giovanni Sambin of lectures given in Padova, June 1980.

[Martin-Löf, 1985] P. Martin-Löf. On the meanings of the logical constants and the justifications of the logical laws. In C. Bernardi and P. Pagli, editors, *Atti degli Incontri di Logica Matematica*, pages 203–281. Univ. di Siena, 1985. Reprinted with bibliographical additions in *Nordic Journal of Philosophical Logic*, 1:11–60, 1996.

[Martin-Löf, 1986] P. Martin-Löf. Unifying Scott's theory of domains for denotational semantics and intuitionistic type theory. In V. Michele Abrusci and Ettore Casari, editors, *Atti del congresso Logica e Filosofia della Scienza, Oggi, San Gimignano, 7–11 Dicembre 1983*, volume I: Logica, pages 79–82. CLUEB, Bologna, 1986.

[Martin-Löf, 1987] P. Martin-Löf. Truth of a proposition, evidence of a judgement, validity of a proof. *Synthese*, 73:407–420, 1987.

[McCarthy, 1960] J. McCarthy. Recursive functions of symbolic expressions and their computation by machine. *Communications of the ACM*, 3:184–195, 1960.

[McCarthy, 1963a] J. McCarthy. A basis for a mathematical theory of computation. In P. Braffort and D. Hirschberg, editors, *Computer Programming and Formal Systems*, pages 33–70. North-Holland Co., Amsterdam, 1963. Earlier version publ. in Proc. of 1961 Western Joint Computer Conference.

[McCarthy, 1963b] J. McCarthy. Towards a mathematical science of computation. In C. M. Popplewell, editor, *Information Processing 62: Proceedings of the IFIP Congress 1962*, pages 21–28. North-Holland Co., Amsterdam, 1963.

[McCarthy, 1981] J. McCarthy. History of LISP. In R. Wexelblat, editor, *History of Programming Languages*, pages 173–197. Academic Press and Association for Computing Machinery (A.C.M.), New York, 1981. Also published in *A.C.M. SIGPLAN Notices* 13 (1978), 217–223.

[McCracken, 1979] N. J. McCracken. *An investigation of a programming language with a polymorphic type structure*. PhD thesis, Syracuse Univ., N.Y., U.S.A., 1979.

[McCracken, 1982] N. J. McCracken. A finitary retract model for the polymorphic lambda-calculus. Technical report, Syracuse Univ., N.Y., U.S.A., 1982. Unpublished.

[Medvedev, 1962] J. T. Medvedev. Finite problems. *Soviet Mathematics Doklady*, 3:227–230, 1962.

[Megill and Bunder, 1996] N. Megill and M. W. Bunder. Weaker D-complete logics. *Journal of the Interest Group on Propositional Logics*, 4:215–225, 1996.

[Menger, 1944] K. Menger. Tri-operational algebra. *Reports of a Mathematical Colloquium*, 5:3–10, 1944. Series publ. by Univ. Notre Dame, Indiana, U.S.A.

[Menger, 1964] K. Menger. On substitutive algebra and its syntax. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 10:81–104, 1964.

[Meredith, 1977] D. Meredith. In memoriam Carew Arthur Meredith. *Notre Dame Journal of Formal Logic*, 18:513–516, 1977.

[Meredith, 1980] D. Meredith. Problems in the analogy between λ-calculus and positive logic. *Bulletin of the European Association for Theoretical Computer Science*, 11:136–137, 1980.

[Meyer and Bunder, 1988] R. K. Meyer and M. W. Bunder. Condensed detachment and combinators. Technical Report TR-ARP-8/88, Research School of the Social Sciences, Australian National Univ., Canberra, Australia, 1988.

[Meyer *et al.*, 1991] R. K. Meyer, M. W. Bunder, and L. Powers. Implementing the "fool's model" of combinatory logic. *Journal of Automated Reasoning*, 7:597–630, 1991.

[Meyer, 1982] A. R. Meyer. What is a model of the lambda calculus? *Information and Control*, 52:87–122, 1982. Preprint circulated 1980. Journal title now *Information and Computation*.

[Mezghiche, 1984] M. Mezghiche. Une nouvelle Cβ-réduction dans la logique combinatoire. *Theoretical Computer Science*, 31:151–164, 1984.

[Milner, 1972] R. Milner. Logic for computable functions; description of a machine implementation. Technical Report STAN-CS-72-288, Stanford Univ., 1972.

[Milner, 1975] R. Milner. Processes: a mathematical model of computing agents. In H. Rose and J. Shepherdson, editors, *Logic Colloquium '73*, pages 157–174. North-Holland Co., Amsterdam, 1975.

[Milner, 1977] R. Milner. Fully abstract models of typed λ-calculi. *Theoretical Computer Science*, 4:1–22, 1977.

[Milner, 1978] R. Milner. A theory of type polymorphism in programming. *Journal of Computer and System Sciences*, 17:348–375, 1978.

[Mints and Tammet, 1991] G. Mints and T. Tammet. Condensed detachment is complete for relevance logic: a computer-aided proof. *Journal of Automated Reasoning*, 7:587–596, 1991.

[Mirimanoff, 1917] D. Mirimanoff. Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles. *L'Enseignement Mathematique*, 19:37–52, 1917.

[Mitschke, 1973] G. Mitschke. Ein algebraischer Beweis für das Church-Rosser Theorem. *Archiv für Mathematische Logik*, 15:146–157, 1973. From author's 1970 PhD thesis, Univ. Bonn, *Eine Algebraische Behandlung von λK-Kalkül und Kombinatorischer Logik*.

[Mitschke, 1979] G. Mitschke. The standardization theorem for the λ-calculus. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 25:29–31, 1979.

[Moldestad, 1977] J Moldestad. *Computations in Higher Types*. Springer-Verlag, Berlin, 1977. Series *Lecture Notes in Mathematics* No. 574.

[Montague, 1974] R. Montague. *Formal Philosophy. Selected Papers*. Yale Univ. Press, New Haven U.S.A. and London, 1974.

[Morris, 1968] J. H. Morris. *Lambda-calculus Models of Programming Languages*. PhD thesis, Massachusetts Inst. Technology, Cambridge, Mass., U.S.A., 1968.

[Moses, 1970] J. Moses. The Function of FUNCTION in LISP, 1970. Massachusetts Institute of Technology, Project MAC report AI–199. Available on Web at ftp://publications.ai.mit.edu/ai-publications/0-499/AIM-199.ps.

[Murthy, 1991] C. Murthy. An evaluation semantics for classical proofs. In *Proceedings Sixth Annual IEEE Symposium on Logic In Computer Science*, pages 96–107. IEEE Computer Society, Los Alamitos, California, U.S.A., 1991.

[Myhill and Shepherdson, 1955] J. Myhill and J. C. Shepherdson. Effective operations on partial recursive functions. *Zeitschrift für Mathematische Logik*, 1:310–317, 1955.

[Nakajima, 1975] R. Nakajima. Infinite normal forms for the λ-calculus. In Böhm [1975], pages 62–82.

[Naur and others, 1963] P. Naur et al. Revised report on the algorithmic language ALGOL 60. *Communications of the ACM*, 6:1–17, 1963.

[Nederpelt and Geuvers, 1994] R. P. Nederpelt and J. H. Geuvers. Twenty-Five Years of Automath Research. In Nederpelt et al. [1994], pages 3–54.

[Nederpelt *et al.*, 1994] R. P. Nederpelt, J. H. Geuvers, and R. C. de Vrijer, editors. *Selected Papers on Automath*. Elsevier, Amsterdam, 1994.

[Nederpelt, 1973] R. Nederpelt. *Strong normalization in a typed lambda calculus with lambda structured types*. PhD thesis, Technical Univ. Eindhoven, Netherlands, 1973. Publ. 1994 in [Nederpelt *et al.*, 1994], pp. 389–468.

[Nerode, 1957] A. Nerode. General topology and partial recursive functionals. Summaries of talks at the Cornell Summer Institute of Symbolic Logic, 1957.

[Nerode, 1959] A. Nerode. Some Stone spaces and recursion theory. *Duke Mathematical Journal*, 26:397–405, 1959.

[Neumann, 1925] J. von Neumann. Eine Axiomatizierung der Mengenlehre. *Journal für die Reine und Angewandte Mathematik*, 154:219–240, 1925. (Correction in *Ibid.* 155, p. 128.) English transl: *An Axiomatization of Set Theory*, in [Heijenoort, 1967], pp. 393–413.

[Neumann, 1928] J. von Neumann. Die Axiomatizierung der Mengenlehre. *Mathematische Zeitschrift*, 27:669–752, 1928. Based on author's doctoral thesis in Hungarian, Univ. Budapest Sept. 1925 (author's footnote p.669). Reprinted in *The Collected Works of J. von Neumann, Vol. 1*, ed. by A. H. Taub, Pergamon Press, Oxford, England 1961, pp.339–422.

[Newman, 1942] M. H. A. Newman. On theories with a combinatorial definition of "equivalence". *Annals of Mathematics*, 43(2):223–243, 1942.

[Newman, 1943] M. H. A. Newman. Stratified systems of logic. *Proceedings of the Cambridge Philosophical Society*, 39:69–83, 1943. Reviewed by A. Church in J. Symbolic Logic 9 (1944), pp. 50–52.

[Nickau, 1994] H. Nickau. Hereditarily sequential functionals. In A. Nerode and Y. V. Matiyasevich, editors, *Proceedings of the Symposium on Logical Foundations of Computer Science*, volume 813 of *Lecture Notes in Computer Science*, pages 253–264. Springer-Verlag, 1994.

[Nivat, 1973] M. Nivat. Langages algébriques sur le magma libre et sémantique des schémas de programme. In M. Nivat, editor, *Automata, Languages and Programming*. North-Holland Co., Amsterdam, 1973. Proc. of 1972 conference.

[Nordström *et al.*, 1990] B. Nordström, K. Petersson, and J. M. Smith. *Programming in Martin-Löf's Type Theory*. Oxford Univ. Press, England, 1990.

[O'Donnell, 1977] M.J. O'Donnell. *Computing in Systems Described by Equations*, volume 58 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1977.

[Ong, 1995] C.-H. L. Ong. Correspondence between operational and denotational semantics. In S. Abramsky, D. Gabbay, and T. S. E. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 4, pages 269–356. Oxford Univ. Press, England, 1995.

[Oosten, 2002] J. van Oosten. Realizability: a historical essay. *Mathematical Structures in Computer Science*, 12:239–263, 2002.

[Paiva, 1989a] V. de Paiva. The Dialectica categories. In J. W. Gray and A. Scedrov, editors, *Categories in Computer Science and Logic*, volume 92 of *Contemporary Mathematics*, pages 47–62. American Math. Society, Providence, R.I., U.S.A., 1989.

[Paiva, 1989b] V. de Paiva. A Dialectica-like model of linear logic. In *Proceedings of the Conference on Category Theory and Computer Science*, volume 389 of *Lecture Notes in Computer Science*, pages 341–356. Springer-Verlag, Berlin, 1989.

[Park, 1976] D. Park. The Y combinator in Scott's lambda calculus models. Theory of Computation Report 13, Dept. of Computer Science, Univ. Warwick, 1976. First version presented 1970 at Symposium on Theory of Programming, Univ. Warwick.

[Peano, 1889] G. Peano. *Arithmetices Principia, Nova Methodo Exposita*. Fratelli Bocca, Torino, Italy, 1889. Also in [Peano, 1958], Vol. 2, Item 16, pp. 20–55. Engl. transls: (1) The Principles of Arithmetic, in *Selected Works of Giuseppe Peano*, ed. H. Kennedy, publ. Allen and Unwin, London, 1973, pp. 101–134; (2) The Principles of Arithmetic, in [Peano, 1958], pp. 83–97.

[Peano, 1895] G. Peano. *Formulaire de Mathématiques*. Fratelli Bocca, Torino, Italy, 1895. Publ. in 5 vols. 1895–1908. Later vols. overlapped and revised earlier vols. Vol. 3 publ. Carré et Naud, Paris, 1901. Vol. 5 titled *Formulario Mathematico* in Peano's language Latino sine flexione.

[Peano, 1958] G. Peano. *Opere Scelte*. Edizioni Cremonese, Milano, 1958. Three volumes.

[Peirce, 1883] C. S. Peirce. The logic of relatives. In C. S. Peirce, editor, *Johns Hopkins Studies in Logic*, pages 187–203. Little, Brown & Co., Boston, U.S.A., 1883. Reprinted in *Collected Papers of Charles Sanders Peirce, Vol. 3, Exact Logic*, ed. by C. Hartshorne & P. Weiss, Harvard Univ. Press, U.S.A., pp. 195–209.

[Peyton Jones, 1987] S. Peyton Jones. *The Implementation of Functional Programming Languages*. Prentice-Hall, England, 1987.

[Phoa, 1990] W. Phoa. Effective domains and intrinsic structure. In *Proceedings Fifth Annual IEEE Symposium on Logic In Computer Science*, pages 366–377. IEEE Computer Society, Los Alamitos, California, U.S.A., 1990.

[Pierce, 2002] B. C. Pierce. *Types and Programming Languages*. MIT Press, Cambridge, Mass., 2002.

[Pigozzi and Salibra, 1998] D. Pigozzi and A. Salibra. Lambda abstraction algebras: coordinatizing models of lambda calculus. *Fundamenta Informaticae*, 33:149–200, 1998.

[Piperno, 1989] A. Piperno. Abstraction problems in combinatory logic; a compositive approach. *Theoretical Computer Science*, 66:27–43, 1989.

[Pitts, 1987] A. M. Pitts. Polymorphism is set theoretic, constructively. In D. H. Pitt, A. Poigné, and D. E. Rydeheard, editors, *Category Theory and Computer Science, Proc. Edinburgh 1987*, volume 283 of *Lecture Notes in Computer Science*, pages 12–39, Berlin, 1987. Springer-Verlag.

[Platek, 1966] R. A. Platek. *Foundations of Recursion Theory.* PhD thesis, Stanford Univ., 1966.

[Plotkin and Abadi, 1993] G. D. Plotkin and M. Abadi. A logic for parametric polymorphism. In M. Bezem and J. Groote, editors, *Typed Lambda Calculi and Applications, TLCA '93*, volume 664 of *Lecture Notes in Computer Science*, pages 361–375. Springer-Verlag, Berlin, 1993.

[Plotkin, 1973] G. D. Plotkin. Lambda-definability and logical relations. Technical Report SAI-RM-4, School of Artificial Intelligence, Univ. Edinburgh, U.K., 1973.

[Plotkin, 1974] G. D. Plotkin. The λ-calculus is ω-incomplete. *Journal of Symbolic Logic*, 39:313–317, 1974.

[Plotkin, 1975] G. D. Plotkin. Call-by-name, call-by-value and the λ-calculus. *Theoretical Computer Science*, 1:125–159, 1975.

[Plotkin, 1977] G. D. Plotkin. LCF considered as a programming language. *Theoretical Computer Science*, 5:223–257, 1977.

[Plotkin, 1978a] G. D. Plotkin. The category of complete partial orders: a tool for making meanings. Lectures, summer school, Dipartimento di Informatica, Università di Pisa, Italy, 1978.

[Plotkin, 1978b] G. D. Plotkin. T^ω as a universal domain. *Journal of Computer and System Sciences*, 17:209–236, 1978.

[Plotkin, 1980] G. D. Plotkin. Lambda-definability in the full type hierarchy. In Hindley and Seldin [1980], pages 363–373.

[Plotkin, 1981] G. D. Plotkin. Post-graduate lecture notes in advanced domain theory. Dept. Computer Science, Univ. Edinburgh, 1981.

[Plotkin, 1993] G. D. Plotkin. Set-theoretical and other elementary models of the λ-calculus. *Theoretical Computer Science*, 121 (Böhm Volume):351–409, 1993. (Part 1 issued 1972 as *A set-theoretical definition of application*, Report MIP-R-95 of School of Artificial Intelligence, Univ. Edinburgh.).

[Plotkin, 1994] G. D. Plotkin. A semantics for static type-inference. *Information and Computation*, 109:256–299, 1994.

[Pottinger, 1978] G. Pottinger. Proofs of the normalization and Church-Rosser theorems for the typed λ-calculus. *Notre Dame Journal of Formal Logic*, 19:445–451, 1978.

[Pottinger, 1980] G. Pottinger. A type assignment for the strongly normalizable λ-terms. In Hindley and Seldin [1980], pages 561–579.

[Prawitz, 1965] D. Prawitz. *Natural Deduction.* Almqvist and Wiksell, Stockholm, 1965.

[Prawitz, 1967] D. Prawitz. Completeness and Haupsatz for second order logic. *Theoria*, 33:246–258, 1967.

[Prawitz, 1968] D. Prawitz. Hauptsatz for higher order logic. *Journal of Symbolic Logic*, 33:452–457, 1968.

[Prawitz, 1971] D. Prawitz. Ideas and results in proof theory. In J. E. Fenstad, editor, *Proceedings of the Second Scandinavian Logic Symposium*, pages 235–307. North-Holland Co., Amsterdam, 1971.

[Quine, 1936a] W. V. Quine. A reinterpretation of Schönfinkel's logical operators. *Bulletin of the American Mathematical Society*, 42:87–89, 1936.

[Quine, 1936b] W. V. Quine. Towards a calculus of concepts. *Journal of Symbolic Logic*, 1:2–25, 1936.

[Quine, 1937] W. V. Quine. New foundations for mathematical logic. *American Mathematical Monthly*, 44:70–80, 1937.

[Quine, 1963] W. V. Quine. *Set Theory and its Logic.* Harvard Univ. Press, Cambridge, Mass., U.S.A., 1963. (Revised edn. 1969).

[Quine, 1972] W. V. Quine. Algebraic logic and predicate functors. In R. Rudner and I. Scheffler, editors, *Logic and Art: Essays in Honor of Nelson Goodman*, pages 214–238. Bobbs Merrill Co., Indianapolis and New York, 1972. (Also publ. 1971 by Bobbs Merrill Co. as a booklet *Algebraic Logic and Predicate Functors*).

[Ramsey, 1926] F. P. Ramsey. The foundations of mathematics. *Proceeding of the London Mathematical Society, Series 2*, 25:338–384, 1926. Reprinted in: *F. P. Ramsey: The Foundations of Mathematics and other Logical Essays*, ed. by R. B. Braithwaite, publ. Routledge and Kegan Paul, New York and London, 1931, pp. 1–61.

[Révész, 1978] G. Révész. λ-calculus without substitution. Technical report, Computer and Automation Institute, Hungarian Academy of Sciences, 1978.

[Révész, 1985] G. Révész. Axioms for the theory of lambda-conversion. *SIAM Journal of Computing*, 14:373–382, 1985.

[Reynolds and Plotkin, 1990] J. C. Reynolds and G. D. Plotkin. On functors expressible in the polymorphic typed lambda calculus. In Huet [1990b], pages 127–152.

[Reynolds, 1972] J. C. Reynolds. Notes on a lattice-theoretic approach to the theory of computation. Report, Syracuse Univ., Syracuse, New York, October 1972. Revised March 1979.

[Reynolds, 1974] J. C. Reynolds. Towards a theory of type structure. In B. Robinet, editor, *Programming Symposium, Proceedings, Colloque sur la Programmation, Paris*, volume 19 of *Lecture Notes in Computer Science*, pages 408–425, Berlin, 1974. Springer-Verlag.

[Reynolds, 1983] J. C. Reynolds. Types, abstraction and parametric polymorphism. In R. Mason, editor, *Information Processing 83*, pages 513–523. North-Holland Co., Amsterdam, 1983.

[Reynolds, 1989] J. C. Reynolds. Syntactic control of interference, part 2. In G. Ausiello, M. Dezani, and S. Ronchi della Rocca, editors, *Automata, Languages and Programming, 16th International Colloquium*, volume 372 of *Lecture Notes in Computer Science*, pages 704–722. Springer-Verlag, Berlin, 1989.

[Reynolds, 1993] J. C. Reynolds. The discoveries of continuations. *Lisp and Symbolic Computation*, 6:233–247, 1993.

[Rezus, 1982] A. Rezus. *A Bibliography of Lambda-Calculi, Combinatory Logics and Related Topics*. Mathematisch Centrum, 413 Kruislaan, Amsterdam, 1982. ISBN 90–6196234–X.

[Rezus, 1986] A. Rezus. Semantics of constructive type theory. *Libertas Mathematica (Arlington, TX)*, 6:1–82, 1986.

[Rimscha, 1980] M. von Rimscha. Mengentheoretische Modelle des λK-Kalküls. *Archiv für Mathematische Logik*, 20:65–74, 1980.

[Robinson, 1965] J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the Association for Computing Machinery*, 12:23–41, 1965.

[Rogers, 1967] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw Hill, New York, 1967.

[Ronchi della Rocca and Venneri, 1984] S. Ronchi della Rocca and B. Venneri. Principal type schemes for an extended type theory. *Theoretical Computer Science*, 28:151–171, 1984.

[Ronchi della Rocca, 1988] S. Ronchi della Rocca. Principal type schemes and unification for intersection type discipline. *Theoretical Computer Science*, 59:181–209, 1988.

[Rosen, 1973] B. K. Rosen. Tree manipulating systems and Church-Rosser theorems. *Journal of the Association for Computing Machinery*, 20:160–187, 1973.

[Rosenbloom, 1950] P. Rosenbloom. *The Elements of Mathematical Logic*. Dover Inc., New York, 1950.

[Rosolini, 1986] G. Rosolini. *Continuity and Effectiveness in Topoi*. PhD thesis, Carnegie Mellon Univ., 1986. Tech. Report CMU-CS-86-123, Dept. of Computer Science, Carnegie Mellon Univ., Pittsburgh, Pennsylvania, U.S.A.

[Rosser, 1935] J. B. Rosser. A mathematical logic without variables, Part 1. *Annals of Mathematics, Series 2*, 36:127–150, 1935. Also Part 2: *Duke Mathematical Journal* 1 (1935), pp. 328–355.

[Rosser, 1942] J. B. Rosser. New sets of postulates for combinatory logics. *Journal of Symbolic Logic*, 7:18–27, 1942.

[Rosser, 1984] J. B. Rosser. Highlights of the history of the lambda calculus. *Annals of the History of Computing*, 6:337–349, 1984.

[Russell and Whitehead, 1913] B. Russell and A. N. Whitehead. *Principia Mathematica*. Cambridge Univ. Press, Cambridge, England, 1913. In 3 parts, appeared 1910–1913. New edn. 1925–1927.

[Russell, 1903] B. Russell. *The Principles of Mathematics*. Cambridge Univ. Press, England, 1903. Also 2nd. edn. 1937, reprinted 1992 by Routledge and Kegan Paul, England.
[Russell, 1919] B. Russell. *Introduction to Mathematical Philosophy*. Allen and Unwin, London, 1919.
[Sallé, 1978] P. Sallé. Une extension de la théorie des types en λ-calcul. In G. Ausiello and C. Böhm, editors, *Automata, Languages and Programming, Fifth Colloquium*, volume 62 of *Lecture Notes in Computer Science*, pages 398–410. Springer-Verlag, Berlin, 1978.
[Sambin, 1987] G. Sambin. Intuitionistic formal spaces: a first communication. In D. Skordev, editor, *Mathematical Logic and its Applications*, pages 187–204. Plenum Press, New York, 1987.
[Sanchis, 1967] L. E. Sanchis. Functionals defined by recursion. *Notre Dame Journal of Formal Logic*, 8:161–174, 1967. Informally circulated 1965.
[Sazonov, 1976a] V. Yu. Sazonov. Degrees of parallelism in computations. In *Mathematical Foundations of Computer Science*, volume 45 of *Lecture Notes in Computer Science*, pages 517–523. Springer-Verlag, Berlin, 1976.
[Sazonov, 1976b] V. Yu. Sazonov. Expressibility of functionals in D. Scott's LCF language. *Algebra i Logika*, 15(3):308–330, 1976.
[Sazonov, 1976c] V. Yu. Sazonov. Functionals computable in series and in parallel. *Sibirskii Matematicheskii Zhurnal*, 17:648–672, 1976.
[Schönfinkel, 1924] M. Schönfinkel. Über die Bausteine der mathematischen Logik. *Mathematische Annalen*, 92:305–316, 1924. English transl: *On the building blocks of mathematical logic*, in [Heijenoort, 1967], pp. 355–366.
[Schröder, 1905] E. Schröder. *Vorlesungen über die Algebra der Logik (Exakte Logik)*. Teubner, Leipzig, Germany, 1905. (In fact publ. in parts. I: 1890, II/1: 1891, III: 1895, II/2: 1905).
[Schroer, 1965] D. E. Schroer. *The Church-Rosser Theorem*. PhD thesis, Cornell Univ., 1965. Informally circulated 1963. 673 pp. Obtainable from University Microfilms Inc., Ann Arbor, Michigan, U.S.A., Publication No. 66-41.
[Schütte, 1960] K. Schütte. Syntactical and semantical properties of simple type theory. *Journal of Symbolic Logic*, 25:305–326, 1960.
[Scott, 1963] D. S. Scott. A System of Functional Abstraction, 1963. Lecture notes, Stanford Univ., informally distributed.
[Scott, 1968] D. S. Scott. An Abstract Theory of Constructions, 1968. Lecture notes, Amsterdam Univ., informally distributed.
[Scott, 1969a] D. S. Scott. A construction of a model for the λ-calculus. Informally distributed, 1969. Notes for a November 1969 seminar, Oxford Univ.
[Scott, 1969b] D. S. Scott. Lattice-theoretic models for the λ-calculus, 1969. Incomplete typescript, 50 pp., Merton College, Oxford Univ.
[Scott, 1969c] D. S. Scott. Models for the λ-calculus. Informally distributed, 1969. Notes, December 1969, Oxford Univ.
[Scott, 1969d] D. S. Scott. A theory of computable functions of higher type. Informally distributed, 1969. Notes for a November 1969 seminar, Oxford Univ.
[Scott, 1969e] D. S. Scott. A type-theoretical alternative to ISWIM, CUCH, OWHY. Informally distributed, 1969. Dated October 1969. Publ. with additions by author in *Theoretical Computer Science* 121 (1–2) (1993), pp. 411–420.
[Scott, 1970a] D. S. Scott. Constructive validity. In M. Laudet, D. Lacombe, L. Nolin, and M. Schützenberger, editors, *Symposium on Automatic Demonstration*, volume 125 of *Lecture Notes in Mathematics*, pages 237–275. Springer-Verlag, Berlin, 1970. Proc. conference in Versailles 1968.
[Scott, 1970b] D. S. Scott. Lattice theory, data types and semantics. In R. Rustin, editor, *Formal Semantics of Programming Languages*, pages 65–106. Prentice-Hall, Englewood Cliffs, N.J., U.S.A., 1970.
[Scott, 1970c] D. S. Scott. Outline of a mathematical theory of computation. In *Proceedings of the Fourth Annual Princeton Conference on Information Sciences and Systems*, pages 169–176. Dept. of Electrical Engineering, Princeton Univ., 1970. Also publ. 1970 as Tech. Monograph PRG-2, Programming Research Group, Oxford Univ., England.
[Scott, 1972] D. S. Scott. Continuous lattices. In F. W. Lawvere, editor, *Toposes, Algebraic Geometry and Logic*, volume 274 of *Lecture Notes in Mathematics*, pages 97–136, Berlin, 1972. Springer-Verlag. Preliminary version publ. 1970 as Tech. Monograph PRG-7, Programming Research Group, Oxford Univ., England.

[Scott, 1973] D. S. Scott. Models for various type-free calculi. In P. Suppes and others, editor, *Logic, Methodology and Philosophy of Science IV*, pages 157–187. North-Holland Co., Amsterdam, 1973. Proc. conference in 1971.

[Scott, 1975] D. S. Scott. Some philosophical issues concerning theories of combinators. In Böhm [1975], pages 346–366.

[Scott, 1976] D. S. Scott. Data types as lattices. *SIAM Journal on Computing*, 5:522–587, 1976.

[Scott, 1977] D. S. Scott. Logic and programming languages. *Communications of the ACM*, 20:634–641, 1977.

[Scott, 1980a] D. S. Scott. Lambda calculus: some models, some philosophy. In J. Barwise et al., editors, *The Kleene Symposium*, pages 223–265. North-Holland Co., Amsterdam, 1980.

[Scott, 1980b] D. S. Scott. Relating theories of the λ-calculus. In Hindley and Seldin [1980], pages 403–450.

[Scott, 1982a] D. S. Scott. Domains for denotational semantics. In M. Nielsen and E. Schmidt, editors, *Automata, Languages and Programming, Ninth International Colloquium*, volume 140 of *Lecture Notes in Computer Science*, pages 577–613. Springer-Verlag, Berlin, 1982.

[Scott, 1982b] D. S. Scott. Lectures on a mathematical theory of computation. In M. Broy and G. Schmidt, editors, *Theoretical Foundations of Programming Methodology*. D. Reidel Co., Dordrecht, Netherlands, 1982. Also publ. 1981 as Tech. Monograph PRG-19, Programming Research Group, Oxford Univ., England.

[Scott, 1986] D. S. Scott. Church's thesis and a unification of type universes. Lecture at conference in Utrecht on Church's Thesis fifty years later, 14–15 June 1986.

[Scott, 1996] D. S. Scott. A new category? Domains, Spaces and Equivalence relations. Unpublished manuscript, December 1996.

[Scott, 2000] D. S. Scott. Some reflections on Strachey and his work. *Higher order and Symbolic Computation*, 13:103–114, 2000.

[Seely, 1989] R. A. G. Seely. Linear logic, $*$-autonomous categories and cofree coalgebras. In J. W. Gray and A. Scedrov, editors, *Categories in Computer Science and Logic*, volume 92 of *Contemporary Mathematics*, pages 371–382. American Mathematical Society, Providence, R.I., 1989.

[Seldin, 1968] J. P. Seldin. *Studies in Illative Combinatory Logic*. PhD thesis, Inst. voor Grondslagenonderzoek, Univ. Amsterdam, 1968.

[Seldin, 1979] J. P. Seldin. Progress report on generalized functionality. *Annals of Mathematical Logic*, 17:29–59, 1979. Condensed from manuscript *Theory of Generalized functionality* informally circulated 1975. Journal now *Annals of Pure and Applied Logic*.

[Seldin, 1980a] J. P. Seldin. Curry's program. In Hindley and Seldin [1980], pages 3–33.

[Seldin, 1980b] J. P. Seldin. A short biography of Haskell B. Curry. In Hindley and Seldin [1980], pages vii–xx. Includes a bibliography.

[Seldin, 1987] J. P. Seldin. MATHESIS: The mathematical foundation of ULYSSES. Technical Report RADC-TR-87-223, Odyssey Research Associates, Rome Air Development Center, Air Force Systems Command, Griffiss Air Force Base, NY 13441-5700, 1987.

[Seldin, 1997] J. P. Seldin. On the proof theory of Coquand's calculus of constructions. *Annals of Pure and Applied Logic*, 83:23–101, 1997.

[Seldin, 2002] J. P. Seldin. Curry's anticipation of the types used in programming languages. *Proceedings of the Canadian Society for History and Philosophy of Mathematics*, 15:148–163, 2002.

[Seldin, 2008] J. P. Seldin. Church and Curry: the lambda calculus and combinatory logic. In D. M. Gabbay and J. Woods, editors, *The Handbook of the History of Logic*, volume 5. Elsevier Co., Amsterdam, 2008.

[Sieg, 1997] W. Sieg. Step by recursive step: Church's analysis of effective calculability. *Bulletin of Symbolic Logic*, 3:154–180, 1997.

[Smyth and Plotkin, 1982] M. B. Smyth and G. D. Plotkin. The category-theoretic solution of recursive domain equations. *S.I.A.M. Journal of Computing*, 11:761–783, 1982.

[Smyth, 1977] M. B. Smyth. Effectively given domains. *Theoretical Computer Science*, 5:257–274, 1977.

[Smyth, 1983] M. B. Smyth. Powerdomains and predicate transformers: a topological view. In J. Diaz, editor, *Automata, Languages and Programming, Tenth International Colloquium*, volume 154 of *Lecture Notes in Computer Science*, pages 662–675. Springer-Verlag, Berlin, 1983.

[Staples, 1981] J. Staples. Efficient combinatory reduction. *Zeitschrift für Mathematische Logik*, 27:391–402, 1981.

[Statman, 1983] R. Statman. λ-definable functionals and $\beta\eta$-conversion. *Archiv für Mathematische Logik*, 23:21–26, 1983.

[Statman, 1985] R. Statman. Logical relations and the typed λ-calculus. *Information and Computation*, 65:85–97, 1985.

[Steele and Gabriel, 1996] G. L. Steele and R. P. Gabriel. The evolution of LISP. In T. Bergin and R. Gibson, editors, *History of Programming Languages – II*, pages 233–330. Addison-Wesley and Association for Computing Machinery (A.C.M.), New York, 1996.

[Steinby and Thomas, 2000] M. Steinby and W. Thomas. Trees and term rewriting in 1910: on a paper by Axel Thue. *Bulletin of the European Association for Theoretical Computer Science*, 72:256–269, 2000. October 2000.

[Stenlund, 1972] S. Stenlund. *Combinators, λ-terms and Proof Theory.* D. Reidel Co., Dordrecht, Netherlands, 1972.

[Stoy, 1977] J. Stoy. *Denotational Semantics: The Scott-Strachey Approach to Programming Language Theory.* M.I.T. Press, Cambridge, Mass., U.S.A., 1977. (Foreword by D. Scott).

[Strachey and Wadsworth, 1974] C. Strachey and C. Wadsworth. Continuations: a mathematical semantics for handling full jumps. Technical Report PRG-11, Oxford Univ. Computing Laboratory, 1974. Reprinted in *Higher-Order and Symbolic Computation* 13:135–152, 2000.

[Strachey, 1967] C. Strachey. Fundamental concepts in programming languages. International Summer School in Computer Programming, Copenhagen, 1967. Reprinted in *Higher Order and Symbolic Computation* 13:11–49, 2000.

[Sundholm, 1986] G. Sundholm. Proof theory and meaning. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Volume III*, pages 471–506. D. Reidel Co., Dordrecht, Netherlands, 1986.

[Szabo, 1974] M. E. Szabo. A categorical equivalence of proofs. *Notre Dame Journal of Formal Logic*, 15:177–191, 1974.

[Szabo, 1978] M. E. Szabo, editor. *Algebra of Proofs.* North-Holland Co., Amsterdam, 1978.

[Tait, 1965] W. W. Tait. Infinitely long terms of transfinite type. In J. Crossley and M. Dummett, editors, *Formal Systems and Recursive Functions*, pages 176–185. North-Holland Co., Amsterdam, 1965. Proc. conference 1963.

[Tait, 1966] W. W. Tait. A nonconstructive proof of Gentzen's Haupsatz for second order predicate logic. *Bulletin of the American Math. Society*, 72:980–983, 1966.

[Tait, 1967] W. W. Tait. Intensional interpretations of functionals of finite type. *Journal of Symbolic Logic*, 32:198–212, 1967.

[Tait, 2001] W. W. Tait. Gödel's unpublished papers on foundations of mathematics. *Philosophiae Mathematica*, 9:87–126, 2001.

[Tait, 2003] W. W. Tait. The completeness of Heyting first-order logic. *Journal of Symbolic Logic*, 68:751–763, 2003.

[Takahashi, 1967] Moto-o Takahashi. A proof of cut-elimination in simple type-theory. *Journal of the Mathematical Society of Japan*, 19:399–410, 1967.

[Takahashi, 1989] Masako Takahashi. Parallel reductions in λ-calculus. *Journal of Symbolic Computation*, 7:113–123, 1989. (Revised version: *Information and Computation* 118 (1995), pp. 120–127).

[Takahashi, 1991] Masako Takahashi. *Theory of Computation, Computability and Lambda-calculus.* Kindai Kagaku Sha, Tokyo, 1991. In Japanese.

[Takeuti, 1953] G. Takeuti. On a generalised logical calculus. *Japanese Journal of Mathematics*, 23:39–96, 1953.

[Tarski and Givant, 1987] A. Tarski and S. Givant. *A Formalization of Set Theory Without Variables*, volume 41 of *Colloquium publications.* American Mathematical Society, 1987.

[Tarski, 1941] A. Tarski. On the calculus of relations. *Journal of Symbolic Logic*, 6:73–89, 1941.

[Tarski, 1953] A. Tarski. A formalization of set theory without variables. *Journal of Symbolic Logic*, 18:189, 1953. (Abstract only).

[Taylor, 1991] P. Taylor. The fixed point property in synthetic domain theory. In *Proceedings Sixth Annual IEEE Symposium on Logic In Computer Science*, pages 152–160. IEEE Computer Society, Los Alamitos, California, U.S.A., 1991.

[Terese, 2003] Terese. *Term Rewriting Systems.* Cambridge Univ. Press, England, 2003. Authors: 'Terese' research group. Editors M. Bezem, J.-W. Klop, R. de Vrijer.

[Thiel, 1995] C. Thiel. Schönfinkel, Moses. In J. Mittelstrass, editor, *Enzyklopädie Philosophie und Wissenschaftstheorie*, pages 726–727. J. B. Metzler, Stuttgart, Germany, 1995.

[Thue, 1910] A. Thue. Die Lösung eines Spezialfalles eines generellen logischen Problems. *Cristiania Videnskabsselskabs Skrifter. I. Mat.-Nat. Kl. 1910*, 8, 1910. Reprinted in *Selected Mathematical Papers of Axel Thue*, ed. by T. Nagel et al., Universitetsforlaget Oslo, 1977, pp. 273–310.

[Tiuryn, 1990] J. Tiuryn. Type inference problems: a survey. In B. Rovan, editor, *Mathematical Foundations of Computer Science 1990*, volume 452 of *Lecture Notes in Computer Science*, pages 105–120. Springer-Verlag, Berlin, 1990.

[Troelstra, 1973] A. S. Troelstra, editor. *Metamathematical Investigations of Intuitionistic Arithmetic and Analysis*, volume 344 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1973. (2nd edn. 1993, publ. as Preprint no. X-93-05 by Institute for Logic, Language and Computation, Univ. Amsterdam, Plantage Muidergracht 24, 1018TV Amsterdam).

[Turing, 1936] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, Series 2*, 42:230–265, 1936. Also see correction in *Ibid.* 43 (1937), pp. 544–546.

[Turing, 1937a] A. M. Turing. Computability and λ-definability. *Journal of Symbolic Logic*, 2:153–163, 1937.

[Turing, 1937b] A. M. Turing. The p-function in λ-K-conversion. *Journal of Symbolic Logic*, 2:164, 1937.

[Turing, 1980] A. M. Turing. Proof that every typed formula has a normal form. Manuscript undated but probably 1941 or '42. Publ. in *An early proof of normalization by A. M. Turing* by R. O. Gandy, in [Hindley and Seldin, 1980], pages 453–455, 1980.

[Turner, 1979] D. A. Turner. A new implementation technique for applicative languages. *Software – Practice and Experience*, 9:31–49, 1979.

[Uspenskii, 1955] V. A. Uspenskii. On enumeration operators. *Doklady Akademii Nauk S.S.S.R.*, 103:773–776, 1955.

[Vickers, 1989] S. Vickers. *Topology via Logic*. Cambridge Univ. Press, England, 1989.

[Vrijer, 1985] R. C. de Vrijer. A direct proof of the finite developments theorem. *Journal of Symbolic Logic*, 50:339–343, 1985.

[Vuillemin, 1974a] J. Vuillemin. Correct and optimal implementations of recursion in a simple programming language. *Journal of Computer and System Science*, 9:332–354, 1974.

[Vuillemin, 1974b] J. Vuillemin. *Syntaxe, sémantique et axiomatique d'un langage de programmation simple*. Thèse de Doctorat d'Etat, Université de Paris VII, Paris, 1974.

[Wadsworth, 1971] C. P. Wadsworth. *Semantics and pragmatics of the lambda-calculus*. PhD thesis, Univ. Oxford, England, 1971.

[Wadsworth, 1976] C. P. Wadsworth. The relation between computational and denotational properties for Scott's D_∞ models of the lambda-calculus. *SIAM Journal of Computing*, 5:488–521, 1976.

[Wadsworth, 1978] C. P. Wadsworth. Approximate reduction and lambda-calculus models. *SIAM Journal of Computing*, 7:337–356, 1978.

[Wadsworth, 1980] C. P. Wadsworth. Some unusual λ-calculus numeral systems. In Hindley and Seldin [1980], pages 215–230.

[Wadsworth, 2000] C. P. Wadsworth. Continuations revisited. *Higher-Order and Symbolic Computation*, 13:131–133, 2000.

[Wand, 1974] M. Wand. On the recursive specification of data types. In E. G. Manes, editor, *Category Theory Applied to Computation and Control*, volume 25 of *Lecture Notes in Computer Science*, pages 214–217. Springer-Verlag, Berlin, 1974.

[Wand, 1977] M. Wand. Fixed-point constructions in order-enriched categories. Technical Report 23, Computer Science Dept., Indiana Univ., Bloomington, Indiana, 1977.

[Wand, 1979] M. Wand. Fixed-point constructions in order-enriched categories. *Theoretical Computer Science*, 8:13–30, 1979.

[Wand, 1987] M. Wand. A simple algorithm and proof for type inference. *Fundamenta Informaticae*, 10:115–122, 1987.

[Welch, 1975] P. H. Welch. Continuous semantics and inside-out reductions. In Böhm [1975], pages 122–146.

[Wells, 1994] J. B. Wells. Typability and type checking in the second-order lambda-calculus are equivalent and undecidable. In *Proceedings Ninth Annual IEEE Symposium on Logic In Computer Science*, pages 176–185. IEEE Computer Society, Los Alamitos, California, U.S.A., 1994.

[Wells, 1999] J. B. Wells. Typability and type checking in System F are equivalent and undecidable. *Annals of Pure and Applied Logic*, 98(1–3):111–156, 1999.

[Wilken and Weiermann, 1997] G. Wilken and A. Weiermann. Sharp upper bounds for the depths of reduction trees of typed λ-calculus with recursors, 1997. Unpublished manuscript.

[Wolfengagen, 2004] V. E. Wolfengagen. *Methods and Means for Computation with Objects*. JurInfoR Ltd., Moscow, Russia, 2004. In Russian.

[Yanovskaya, 1948] S. A. Yanovskaya. Osnovaniya matematiki i matematicheskaya logika. In *Matematika v SSSR za Tridkat let 1917–1947*, pages 9–50. OGIZ, Moscow, 1948. In Russian. Reviewed by G. L. Kline in *J. Symbolic Logic* 16 (1951), pp.46–48.

THE LOGIC OF CHURCH AND CURRY

Jonathan P. Seldin

In the ordinary set-theoretic foundations of mathematics, functions are defined to be sets of ordered pairs in which no two distinct pairs have the same first element. Then if x is in the domain of a function f, $f(x)$ is defined to be the second element of the ordered pair whose first element is x. Thus, for example, the squaring function on the real numbers would be defined as

$$f = \{(x, x^2) : x \text{ is a real number}\},$$

and $f(3) = 3^2 = 9$.

However, there is another way to think of functions: as rules of correspondence. Thinking of the squaring function on the real numbers this way, we might represent it using the notation

$$x \mapsto x^2,$$

and the evaluation of the function at $x = 3$ would be written

$$(x \mapsto x^2)(3) = 3^2 = 9.$$

In fact, this idea occurred in the late 1920s to Alonzo Church, although instead of $x \mapsto x^2$ he wrote

$$\lambda x \,.\, x^2,$$

and for the evaluation of the function at $x = 3$ he wrote

$$(\lambda x \,.\, x^2)3 = 3^2 = 9.^1$$

This approach to functions is the basis of the work of both Alonzo Church (1903-1995) and Haskell Brooks Curry (1900-1982). Both of them set out to base logic and mathematics on functions instead of on set theory, and although neither of them achieved the kind of success he first sought, both of them wound up making major contributions to the foundations of logic and mathematics. Furthermore, the results of their work have come to be extremely important in theoretical computer

[1] This was not his original notation; his original notation was $\{\lambda x[x^2]\}(3) = 3^2 = 9$; see [Church, 1932]. But the notation given in the text had become standard by 1941 [Church, 1941]. By the way, Church did not write "=" here, but wrote "conv" for *conversion*. In this paper, I will write $=_*$ for an unspecified conversion relation.

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

science. The system introduced by Church is *λ-calculus*, the system of Curry is *combinatory logic*.[2]

For the convenience of readers, a complete description of lambda-calculus and combinatory logic is given in the Appendix of this article.

A history of both lambda-calculus and combinatory logic appears in this volume [Cardone and Hindley, 2008]. To avoid unnecessary duplication, I will follow a systematic rather than a historical order. Although combinatory logic began earlier than λ-calculus, it is easier to understand combinatory logic if one already knows λ-calculus. Therefore, I will begin with the latter.

I would like to thank Roger Hindley and Martin Bunder for their helpful comments and suggestions.

1 LAMBDA-CALCULUS

1.1 Background of Alonzo Church

Alonzo[3] Church was born on June 14, 1903 in Washington D. C. He received his A.B. at Princeton University in 1924. In later life, he reported that he was drawn to things of a fundamental nature, and while he was an undergraduate he published a minor paper about the Lorentz transformation, the foundation of the special theory of relativity [Church, 1924]. He also caught the eye of Oswald Veblen, who made important contributions to projective and differential geometry and topology, and who was interested in foundational issues in the sense of postulate systems for geometry. After Church graduated, the Department of Mathematics at Princeton gave him a fellowship to go to graduate school. While he was a graduate student, he published another paper related to foundational matters, [Church, 1925]. He completed his Ph.D. in 1927 with a dissertation on systems in which the axiom of choice might be false [Church, 1927].

After finishing his Ph.D., Church spent two years on a National Research Fellowship, one year at Harvard and another year at Göttingen and Amsterdam. Then, in 1929, he returned to Princeton to take up a faculty position in the Department of Mathematics. During this period, he published [Church, 1928], which was related to foundational matters, though not to those which later led him to introduce lambda-calculus.

---

[2]Curry came to use the name "combinatory logic" to refer to both his system, which he called *synthetic combinatory logic*, and λ-calculus, but most people now use the name only for a variant of Curry's system.

[3]The material in this section comes from [Enderton, 1995] and from the transcript of an interview of Church by William Aspray on 17 May 1985, available on the web at `http://libweb.princeton.edu/libraries/firestone/rbsc/finding_aids/mathoral/pmc05.htm`.

1.2 Origins of Lambda-Calculus

Until he completed his Ph.D., Church's approach to foundational matters in the above mentioned papers been informal, but while he held his National Research Fellowship, in 1928–29, Church began to develop a system of formal logic that he hoped would be adequate for the foundations of mathematics. He wanted to build a type-free system in which all variables occurred only bound, and he hoped to avoid contradiction by restricting the law of excluded middle.[4] He included in this system notations for function application and abstraction, which had previously appeared in the work of Peano and Frege [Cardone and Hindley, 2008, §2], but apparently Church was not then familiar with these works [Cardone and Hindley, 2008, §4.1]. However, unlike Peano and Frege, Church gave formal rules for them and began to analyze their consequences in depth. This system was published in [Church, 1932]. His atomic constants were all logical operators, Π for a relative universality,[5] Σ for existence, & for conjunction, $\sim$ for negation, ι for the definite article, and A for a kind of abstraction operator. Well-formed formulas were formed from these atomic constants and variables by means of application and λ-abstraction, where in the latter, $\lambda x\ .\ M$ was well-formed whether or not x occurred free in M. In addition to rules for the logical constants and certain axioms, he had three rules of procedure for abstraction, application, and substitution: (1) change of bound variables (α-conversion, using modern terminology); (2) β-contraction, where, in the redex $(\lambda x\ .\ M)N$, x must occur free in M; (3) the reverse of 2. There were also two rules for logical operators:[6] (4) for well formed M and N,

$$MN \vdash \Sigma M,$$

and (5) for well formed M, N, and P,

$$\Pi MN, MP \vdash NP.$$

The paper concluded with proofs of a combination of the deduction and generalization theorems for Π:

1. Let x be a variable occurring free in M, and let $\vdash \Sigma(\lambda x\ .\ M)$, and let $M \vdash N$. Then if x occurs free in N, then $\vdash \Pi(\lambda x\ .\ M)(\lambda x\ .\ N)$. If x does not occur free in N, then $\vdash N$.

[4]The idea was that without the law of excluded middle, the existence of the Russell class R would not imply a contradiction, since it would not be true that R is the sort of thing which can be an element of R, and without the law of excluded middle it would be impossible to prove $R \in R \vee R \notin R$. In Church's system, as in Curry's, this paradox would be represented not in terms of classes, but in terms of predicates, so that instead of the Russell class R, there would be a Russell predicate R, such that RR is equivalent to $\sim(RR)$. Then, if R is not automatically in the domain of R, one would need $RR \vee \sim(RR)$ to derive the paradox.

[5]ΠMN is to be distinguished from $(\forall x)(Mx \supset Nx)$. The latter requires that MP and NP be defined for all possible values of P, whereas the former only requires that NP be defined and true whenever MP is true, so that NP may be undefined if MP is defined and not true.

[6]Stated here in modern notation.

2. If $\vdash \Sigma M$ and x does not occur free in N and $Mx \vdash Nx$, then $\vdash \Pi MN$.

3. If $\vdash \Sigma M$ and x does not occur free in N and $Mx \vdash N$, then $\vdash N$.

Church did not have in this paper a general notion of deduction from premises, only deduction from one premise.

Very soon after writing [Church, 1932], Church discovered that his list of axioms was contradictory, so in [Church, 1933] he modified it to avoid that contradiction, discussed the possibility of a consistency proof, derived a large number of results in the system, discussed the term that represents the Russell paradox, and ended with his proposal, now well-known, for representing the natural numbers, here given in modern notation: see Remarks 9 and 10 in Appendix A. In terms of these definitions, Church defined

$$\mathcal{N} \equiv \lambda x \,.\, (\forall P)(P\mathbf{1} \wedge (\forall u)(Pu \supset P(\boldsymbol{\sigma} u)) \supset Px).$$

Church noted that the first two Peano axioms and the induction axiom follow immediately from these definitions.

By this time, Church had two graduate students who later became two of the leaders of mathematical logic, Steven Cole Kleene (1909–1994) and John Barkley Rosser (1907–1989). Church set Kleene the problem of studying his system, especially the theory of natural numbers. Both worked on matters connected with Church's formal logic: Kleene, who started first, worked on the system itself, while Rosser worked on its connection to Curry's combinatory logic.

The first result of these studies was [Kleene, 1934], which showed that arguments using the intuitive rule of proof by cases could be carried out in that system. This made possible definition by cases, which Kleene needed for his dissertation. Kleene modified Church's system in relatively minor ways, the most important being that $\lambda x \,.\, M$ was to be well-formed only when x occurred free in M. This modification was retained in all later publications having to do with untyped λ-calculus by Church, Kleene, and Rosser. In this paper Kleene also published for the first time a definition of combination by defining I to be $\lambda x \,.\, x$ and, following Rosser, J to be $\lambda uxyz \,.\, ux(uzy)$ and then taking combinations to be terms formed from I and J by applications. Kleene then reproduced a result from Rosser's as yet unpublished dissertation that for any term a combination could be found to which the term converts. He needed this result to show that definitions by cases and proofs by cases could be carried out in Church's system. This was the first published result relating λ-calculus and combinatory logic. (Rosser's dissertation was later published as [Rosser, 1935a; Rosser, 1935b]; see § 2.3 below.)

Next came Kleene's dissertation [Kleene, 1935], which was based on Church's definition of natural numbers. Kleene showed that Church's system was adequate for the elementary theory of natural numbers by proving the third and fourth Peano axioms. In the course of doing this, he showed that a number of (recursive) functions could be represented as λ-terms, including functions such as the predecessor which Church had not thought could be represented that way. Rosser had also found some such representations in work that was never published, and by

the fall of 1933, Church began to speculate that every function of positive integers that can be effectively calculated can be represented as a λ-term. By early the following year, Church was convinced of what is now known as "Church's Thesis." See [Rosser, 1984, p. 334].

But meanwhile, in late 1934, Kleene and Rosser succeeded in proving that Church's system of logic was inconsistent by showing that Richard's paradox could be derived in it. [Kleene and Rosser, 1935]. Their proof also applied to the system of combinatory logic then being developed by Curry (see §2.3 below).

Church, Kleene, and Rosser responded to this inconsistency by extracting from Church's system of logic what is now called the pure λ-calculus: the part of the system involving the formation of λ-terms and the rules for reduction and conversion. In Church's Thesis they already had an indication of the importance of pure λ-calculus, and in the years following the discovery of the inconsistency, they obtained a number of major results, results which fully justified treating the pure λ-calculus as a system on its own.

1.3 First Important Results using Pure $\boldsymbol{\lambda}$-Calculus

The first result needed to justify the pure λ-calculus was a consistency proof, and one was published as [Church, 1935]. This result was actually proved for an extension of the pure λ-calculus which included an axiom, **2** (here taken as the truth value for "true"), and a new operator δ with axioms to ensure that $\delta MN = \mathbf{2}$ if M and N are in normal form and they differ only by changes of bound variables and $\delta MN = \mathbf{1}$ (here taken as the truth value for "false") if M and N are in normal form and are not the same except for changes of bound variables. Church's intention was to incorporate logic by using **2** and **1** as the truth values, and he defined the logical connectives for negation and conjunction simply as functions on natural numbers and then defined the logical quantifiers in such a way that when a quantifier formula is true it converts to **2** but otherwise has no normal form. The consistency result depended on some results that Church obtained with Rosser and which were published soon after in [Church and Rosser, 1936]. This latter paper proved what is now known as the "Church-Rosser Theorem" or "confluence theorem," which is now considered basic in term rewriting. It says that if $M \rhd P$ and $M \rhd Q$, then there is a term N such that $P \rhd N$ and $Q \rhd N$. See Theorem 11 in Appendix A below.

This was followed by a result of Kleene showing that the functions that can be represented as λ-terms are precisely those which are recursive in the Herbrand-Gödel sense [Kleene, 1936b]. This gave important support to Church's Thesis, which received additional support a couple of years later by work of Alan Turing [1936–1937; 1937a; 1937b].

But the most important result derived at this time from pure λ-calculus, the result for which Church is most famous, is his proof that there are unsolvable problems [Church, 1936c]. In this paper, Church proved that there is no effectively computable procedure (in the form of a recursive function) which will decide

whether or not two given terms convert to each other, and he also proved that there is no effectively computable procedure that will determine whether or not a term has a normal form. A Gödel numbering converts these problems into problems of elementary arithmetic, and Church ended his paper by pointing out that his result implies that any formal system which includes elementary arithmetic and is ω-consistent is undecidable. In [Church, 1936b; Church, 1936a], Church extended this result to the undecidability of first-order logic. Rosser, in his [1936], showed that the assumption of ω-consistency is unnecessary, both for Gödel's incompleteness theorem and Church's undecidability result.

There was one more use of λ-calculus made in the 1930s by Church and Kleene, who showed in their [1936] how to represent some ordinal numbers as λ-terms. Church discussed this representation and its relation to constructivity in his [1938].

A summary of all results on λ-calculus up to this point appears in [Church, 1941].

After these results were obtained, Church, Kleene, and Rosser gave up on basing a system of logic on λ-calculus. Kleene moved on to recursive function theory, starting with [Kleene, 1936a]. When he wrote further on ordinal notations, he did not deal with representations in λ-calculus [Kleene, 1938; Kleene, 1944; Kleene, 1955].[7] Rosser's work on combinatory logic will be taken up in §2.3 below. And Church himself gave up on trying to construct a type-free logic based on λ-calculus.

1.4 Church's Type Theory

Although he had given up on basing type-free logic on λ-calculus, he did use λ-calculus as a basis for a system of simple type theory [Church, 1940]. In this system, the types were defined as in Definition 4 of Appendix B, where the atomic types were o for propositions and ι for individuals.[8] Terms were allowed in the system only if they could be assigned types by the typing rules of TACh_λ of Definition 5 of Appendix B. The atomic terms included terms for the boolean logical operators negation and disjunction and also a universal quantifier for every type and, in addition, a definite descriptions operator $\iota_{\alpha(o\alpha)}$ of type $(\alpha \to o) \to \alpha$ for every type α. The axioms and rules were sufficient to give higher-order classical logic with equality and descriptions, and there was an axiom which asserted the existence of two distinct individuals. In the paper, Church proved the Peano postulates using the typed version of his representation of the natural numbers, and he also proved that the definition of functions by primitive recursion could be carried out in the system.

A few years later, Church [1951] used this type theory in formulating a logic of sense and denotation. And still later, he applied this type theory in a number of papers including his [1973; 1974; 1989; 1993].

[7] Kleene says in his [1981] that he stopped using λ-calculus as a basis for his work because of the negative reactions of audiences to his early papers which did use λ-calculus. The only later paper he wrote that was based in λ-calculus is [Kleene, 1962].

[8] But Church wrote '$(\beta\alpha)$' for $\alpha \to \beta$.

Otherwise, he published nothing further involving λ-calculus.

Leon Henkin [1950] proved Church's type theory complete. In his [1963], he gave a formulation in which the only primitive constants are the identity relations over each type. Peter B. Andrews extended it to include transfinite types in his [1965], and more recently he has included it in an introductory textbook, his [1986].

This kind of system has been extended in ways important for theorem provers and proof development systems. The first such extensions were due to Per Martin-Löf [1975; 1984] and Jean-Yves Girard [1971; 1972]. Martin-Löf's type theory was used by Robert Constable and his team as the basis for the Nuprl proof development system [Constable and others, 1986], while the systems of Girard were extended by Coquand and Huet to the calculus of constructions [Coquand and Huet, 1988], which formed the basis for the Coq proof assistant.[9]

For a general overview of typed systems see [Kamareddine *et al.*, 2004].

2 COMBINATORY LOGIC

2.1 Background of Haskell Brooks Curry

Haskell[10] Brooks Curry was born on September 12, 1900 at Millis, Massachusetts. His parents ran the School for Expression, now known as Curry College. Curry entered Harvard University as an undergraduate in 1916, and originally intended to study medicine. However, when the United States entered World War I, he changed his major to mathematics. His original motivation was patriotic: he thought knowing mathematics would enable him to serve his country in an artillery unit. After his eighteenth birthday, he joined the Student Army Training Corps, but before he saw any action, the war ended.

Curry graduated from Harvard with the degree of A.B. in 1920 and began his graduate work in a program in electrical engineering at the Massachusetts Institute of Technology that involved working half-time at the General Electric Company. However, he became dissatisfied with the engineers because they did not share his interest in why results were true, so after two years he returned to Harvard University as a student of physics. For the first year he was in that program, 1922–23, he was a half-time research assistant to P. W. Bridgeman. He received his M.A. in physics in 1924. He then went back to mathematics, where he entered a Ph.D. program and was eventually assigned by George D. Birkhoff a dissertation topic on differential equations. However, he was becoming increasingly interested in logic, and this led him eventually to combinatory logic.

[9] See `http://coq.inria.fr/`.

[10] Much of the information in this section and those which follow comes from my personal association with Curry. I was his research assistant at the Pennsylvania State University from 1964–1966, helped him train new assistants at the University of Amsterdam in 1966–67, finished my doctorate under his direction in 1968, and then became a co-author. I remained in touch with him for the rest of his life.

2.2 Origins of combinatory logic

Curry's first look at a work on logic came in May of 1922, when he first studied [Whitehead and Russell, 1910–1913]. He immediately noticed that of the two rules of inference in Chapter 1 of that work, the rule of substitution is substantially more complicated than the rule of modus ponens.[11] He gradually began to feel a desire to break down the rule of substitution into simpler rules, and during 1926–1927 this brought him to the combinators I, B, C, and W[12] and the application operation. At that time it was standard to make underlying assumptions about substitution as part of what Curry called the "prelogic" of a system, not stated explicitly. Curry felt that it was important to write these assumptions out formally, so that they could be analyzed. At this point, he decided that he wanted to work in logic, and so he abandoned his work in differential equations. However, there was nobody at Harvard who could supervise a dissertation on this topic.

As a result, for the year 1927–1928, Curry took an instructorship at Princeton University.[13] During that year, he discovered in a library search that his work had been anticipated by Moses Schönfinkel [1924]. Schönfinkel had shown that all the combinators could be defined in terms of K and S, and Curry immediately saw the point of K and incorporated it in his system along with I, B, C, and W. However, for some years, Curry did not see the point of replacing I, B, C, and W by S. Curry also noticed that Schönfinkel had taken equality (conversion) informally, without giving any axioms or rules for it. Curry realized that despite Schönfinkel's work, there was scope for a dissertation on this subject.

For this purpose, Curry sought the advice of Oswald Veblen. Veblen and Curry then sought information about Schönfinkel from the topologist Paul Alexandroff, who was visiting Princeton that year. Alexandroff informed them that Schönfinkel had returned to the USSR and was in a mental instution, so there was no chance that Curry could work with him. However, Schönfinkel's paper was based on a talk he had given to the Göttingen Mathematical Society in December of 1920, and there were people there who knew about this subject: Heinrich Behmann and Paul Bernays.

So Curry decided to go to Göttingen to get his doctorate. He wrote his first paper, [Curry, 1929], as part of an application for a grant to make that trip. In this paper, Curry followed Schönfinkel in taking K and S as basic combinators, application as the operation that formed terms, and $=_*$ (which Curry then wrote as '=' and thought of as equality) as the only other primitive, but unlike Schönfinkel he gave formal axioms and rules for his system. He called attention to the fact that the reason for the interest in this system is its *combinatory completeness*, which

[11] The notes Curry made at the time can be seen on `http://www.sadl.uleth.ca` under the title "Works of Haskell Curry." Search by date for the item of 20 May 1922.

[12] See Appendix C for the definitions. Curry's original notation for these combinators was I, B, C, and W. Curry first used the sans-serif notation for combinators and other constants in [Curry and Feys, 1958].

[13] He just missed Alonzo Church, who had completed his Ph.D. and left to study elsewhere for two years. See § 1.1 above.

says that if X is any term in which some of the variables $x_1, x_2, \ldots, x_n$ occur, then there is a term A, in which none of these variables occur, such that

$$Ax_1x_2 \ldots x_n =_* X.$$

Curry wanted this A to be uniquely determined by X and the variables $x_1, x_2, \ldots, x_n$, so one of his rules was a generalization of rule (ζ)[14]: if

$$Xx_1x_2 \ldots x_n =_* Yx_1x_2 \ldots x_n$$

for variables $x_1, x_2, \ldots, x_n$ which do not occur in X or Y, then

$$X =_* Y.$$

However, this rule was too complicated from the point of view of analyzing the logical rule of substitution into simpler rules, so Curry wanted to find a set of postulates which would imply this rule. In [Curry, 1929], he gave axioms that imply this rule in a special case, which he predicted would be important for the analysis of substitution.

Curry spent the year 1928–1929 in Göttingen completing his dissertation [Curry, 1930]. His official supervisor was David Hilbert, but he actually worked with Bernays. Church was in Göttingen for half of that year, and in his later years Curry remembered seeing a manuscript of Church's with many occurrences of the symbol 'λ', but at the time he had no idea that this manuscript had any connection with what he was doing. He did not realize there was a connection until he saw [Church, 1932] a few years later.

The system of [Curry, 1930] is based on the combinators B, C, W, and K, and had addional atomic terms Q (for logical identity, which here meant conversion), Π (for the universal quantifier), P (for implication), and Λ (for conjunction). The term forming operation was application, and formulas were of the form '$\vdash X$' for 'X is provable.' The axioms included

$$\vdash \Pi(\mathsf{W}(\mathsf{CQ})),$$

from which the reflexive law of Q can be proved (using the rules given below) and axioms from which the generalization of rule (ζ) follows. For example, the axiom giving the conversion rule for K is

$$\vdash \mathsf{Q}(\mathsf{K}XY)X.$$

The rules included rules from which it can be proved that Q has all the properties of equality and weak conversion, and the following logical rules:

Rule Π. $\Pi X \vdash XY$,

Rule P. $\mathsf{P}XY, X \vdash Y$,

[14]See Appendix A, Remark 1.

Rule Λ. $X, Y \vdash \Lambda XY$.

For this system, Curry proved combinatory completeness by giving an algorithm for what he later called *bracket abstraction*, although at this time he did not use that notation, and he described his algorithm as defining, for a combination X in which some of the variables $x_1, x_2, \ldots, x_n$ occur, a term A in which none of those variables occur such that

$$Ax_1x_2 \ldots x_n =_* X.$$

The first step is to remove the variables from $x_1, x_2, \ldots, x_n$ which do not occur in X using K and B. The second step is to repeat variables as often as they occur in X using W and B. Third, the variables must be arranged in the order in which they occur in X using C and B. And finally, parentheses that occur in X are inserted using B and I.

Curry also gave a consistency proof for the system he had defined. However, no logical axioms for Π, P, or Λ were given, and the development of the logical part of the system was left for the future.

2.3 Early Development of Combinatory Logic

After finishing his dissertation, Curry joined the faculty of the Pennsylvania State College in September 1929. At first, he was not happy there; he had been at Harvard and Princeton, and Penn State did not then support research,[15] he felt that Penn State was an institution of much lower standing. However, his choice of logic as a field had cut down his opportunities, and the Great Depression began that fall, so finding a job elsewhere was out of the question.

Since Penn State had heavy teaching loads, Curry did not have much time for research there for the next three years. There was also no support for graduate students at that time. In fact, Curry's first graduate student, Edward Cogan, did not start working with him until 1950. So, unlike Church, Curry had to continue his development of combinatory logic on his own.

However, he did manage to do some research during his first years. During the very first year of teaching at Penn State, he published [Curry, 1931], in which he added postulates for the universal quantifier and proved some of its properties. But his further research had to wait until he was able to take a year off from teaching as the result of a National Research Fellowship, which he started in 1931. The fellowship was originally supposed to be for two years, from 1931 to 1933. However, after one year, his fellowship for the second year was cancelled because he had a job to go back to, and many others did not. Nevertheless, this one year, which he spent at the University of Chicago, was fruitful for him. He published [Curry, 1932; Curry, 1933; Curry, 1934a; Curry, 1934b; Curry, 1936] and the abstract [Curry, 1935], all of which were written or partly written during this year.

[15] It is very different now.

Of these, [Curry, 1932] contained a few technical modifications of [Curry, 1930], [Curry, 1933] introduced the notation of bracket abstraction, which allowed statements using bound variables to occur as abbreviations, and [Curry, 1934b] extended the logic by introducing axioms for equality and implication; it included a proof of the Deduction Theorem (equivalent to implication introduction in natural deduction) and the following generalization of it [Curry, 1934b, Theorem 14]:[16] *if*

$$A_1, A_2, \ldots, A_m \vdash X,$$

then there is a proof in which the variables $x_1, x_2, \ldots, x_n$ *do not occur of*

$$\vdash (\forall x_1)(\forall x_2)\ldots(\forall x_n)(A_1 \supset (A_2 \supset (\ldots(A_m \supset X)\ldots))).$$

The paper [Curry, 1934a] is an extended abstract of [Curry, 1936], which Curry had some trouble getting accepted. In these papers, Curry proposed a theory of predication or *functionality*, which would now be called a theory of categories or types.[17] These types are predicates, so what is now usually written as $X : \alpha$ was written by Curry as αX, so that the type, α, is the predicate and the term, X, is the subject. This is the source of the name of the Subject-Reduction Theorem. The main operator of this theory is F, which is characterized by the rule $\mathsf{F}\alpha\beta X, \alpha Y \vdash \beta(XY)$, so $\mathsf{F}\alpha\beta X$ is essentially equivalent to $X : \alpha \to \beta$.[18] This means that $\mathsf{F}\alpha\beta$ is the type of functions which map arguments of type α into values in type β. (This does not necessarily mean that α is the domain of such functions, only that this domain includes α.) Furthermore, subjects and predicates were not completely separated in the original theory. In [Curry and Feys, 1958, §10A3], Curry derives $\mathsf{K}X(\mathsf{WWW})$ for an arbitrary term X, and then applies a reduction step to deduce X, thus proving the full free system of [Curry and Feys, 1958, §10A3] inconsistent.

Curry had the original idea for functionality in December 1928.[19] Interestingly, his first notation for $\mathsf{F}\alpha\beta$ was $F_{\alpha\beta}$, but he found this awkward, and so tried the current notation, $[\alpha \to \beta]$, before rejecting that and settling on $F\alpha\beta$.[20] Among the categories Curry wanted in his system is H (originally called Pr), standing for *proposition*. Curry intended to use the theory of functionality to restrict logical postulates to some appropriately defined (by logical postulates) category of propositions as a means of avoiding any paradoxes.[21]

[16] Stated here using modern abbreviations.

[17] However, the theory of functionality was originally much less restrictive than type theory. It *mentioned* types, but it did not use them to define what it meant for a formula to be "well-formed," so that Curry originally anticipated allowing untyped terms in the system. This made it much less restrictive than Church's simple theory of types [Church, 1940].

[18] For Curry, this was roughly equivalent to $(\forall x)(\alpha x \supset \beta(Xx))$, but differed from that the way Church's ΠXY differed from $(\forall x)(Xx \supset Yx)$. See Footnote 5.

[19] See T281213A, the first item for December 13, 1928, under Works of Haskell Curry at `http://www.sadl.uleth.ca`.

[20] This became $\mathsf{F}\alpha\beta$ in [Curry and Feys, 1958]. See also [Seldin, 2003].

[21] Curry observed in these papers that there is a term, $[x](\neg(xx))$, which represents Russell's

Meanwhile, Rosser published his dissertation [Rosser, 1935a; Rosser, 1935b], in which he presented a system of combinatory logic based on the combinators I and J. Rosser's rules for combinator conversion were as follows:

$$\begin{aligned} \mathsf{I}X &\Leftrightarrow X, \\ W(\mathsf{I}X) &\Leftrightarrow WX, \\ W(\mathsf{J}XYZU) &\Leftrightarrow W(XY(XUZ)), \end{aligned}$$

where W is an arbitrary term.

One of Rosser's aims was to eliminate any primitive rules whose conclusions are not uniquely determined by the premises. One such rule is Curry's rule for Π, which is $\Pi X \vdash XY$, since the choice of Y is arbitrary here. Rosser avoided this problem by using Church's Π, which is characterized by the rule

$$\Pi XY, XZ \vdash YZ.$$

To get the effect of a universal quantifier, he followed Church in using ΠE, where E is Church's E, which is defined by

$$E \equiv \lambda x \,.\, (\exists X)(Xx).$$

2.4 The Paradox of Kleene and Rosser and Curry's New Program for Combinatory Logic

Curry's papers up to this time continued the development of combinatory logic along the lines he had previously set out, and the abstract [Curry, 1935] had taken him up to the representation of set theory in combinatory logic. He seemed on the point of showing that combinatory logic could serve as a foundation for logic and mathematics.

However, Kleene and Rosser proved that Curry's system of [Curry, 1934b], like that of Church, was inconsistent. Their result appeared in print in [Kleene and Rosser, 1935], but Curry knew about it before that. He had been in correspondence with Rosser, who had informed him in December, 1933[22] that he thought he could prove Church's system inconsistent. Curry mentioned in a reply dated December 18, 1933 that he would like to see the proof of the contradiction Rosser mentioned as soon as possible. In a letter to Curry postmarked March 2, 1934, Rosser said that he and Kleene had finished the first draft of a proof that Church's system is inconsistent. Then in a joint letter dated November 15, 1934, Kleene and Rosser sent Curry a manuscript of [Kleene and Rosser, 1935]. The text of their letter began

paradox (stated in terms of predicates in stead of sets), and he proposed to find postulates from which it could not be proved that the application of this term to itself is a proposition. However, it was important for Curry that this term exist in his system.

[22]Rosser did not date the letter, so I am relying on a postmark which is smudged. I think the date is about December 10, but it might be a few days later.

> We are sorry to say that we have a proof of the inconsistency of your system of formal logic. We are also sorry that we did not obtain our results in definite form in time to let you know sooner.

Curry's reply to both of them, dated November 19, 1934, began

> Your letter of November 15th enclosing reprint and manuscript has just been received. You begin by saying that you are sorry that you have done this dreadful thing. I am afraid that I can not share your grief. In the first place I do not believe a word of it;– you are not sorry but you are full of that profound satisfaction which comes from doing something worth while. In the second place I have no occasion to feel sorry for myself; for I feel that whatever advances or deepens our knowledge of the subject to which I have devoted so much attention advances my interest as well. Consequently I am not inclined to feel sorry but just the opposite, and I hasten to congratulate you both.

A little later, he said

> . . . you seem to imply that you have refuted my position by showing that my system as you call it is inconsistent. On the contrary your results are in general agreement with my position and are in a sense a continuation of my work. You allude for example to my paper [reference to [Curry, 1934b]]. I proved in that paper (Theorem 14) that any logical system of a certain type, which I called an L system had a certain property, say T. As I understand it you have essentially proved that any system having the property T is "inconsistent". Very well, that simply means that any L system is inconsistent, which is a very much stronger result that (*sic*) which I had proved. Whether or not you needed my result in order to establish the stronger one I do not know; but in any case your result does not contradict mine. The conclusion is that a consistent system of logic, if any, must be weaker than an L system which is a positive result of some value.

Curry then went on to point out that the contradiction requires some of the postulates of his [1934b], and that the systems of his earlier papers, which did not include these assumptions, might well be consistent.

Thus, whereas Church, Kleene, and Rosser reacted to the inconsistency by abandoning the attempt to find a type-free formal logic that could serve as a basis for mathematics, Curry did not. He had already developed a strategy for dealing with contradictions, namely using a predicate to stand for "proposition" together with his theory of functionality to restrict the postulates for the logical operators. The contradiction of Kleene and Rosser only meant for him that he was required to face this problem earlier than he had anticipated in the development of the subject.

The first decision Curry made on learning of the paradox was that he needed to study it in detail "so as to lay bare its central nerve."[23] The direct result of this study was [Curry, 1941b], which shows how to derive Richard's paradox in a very general setting. Meanwhile, he also published [Curry, 1941c], in which he revised the fundamental postulates of combinatory logic taking into account some of the ideas of Rosser's thesis [Rosser, 1935a; Rosser, 1935b]. This was closely followed by [Curry, 1941a], in which he proved the consistency of a system stronger than that of [Curry, 1930]. Finally, he found a much simpler contradiction that could be derived under the assumptions of the Kleene-Rosser paradox in [Curry, 1942b]. This simpler contradiction, known as *Curry's paradox*, can be stated easily as follows: Suppose the system has an operator P for implication, so that $X \supset Y$ is an abbreviation for $\mathsf{P}XY$. Suppose that the following rules hold, either as primitive rules or as derived rules:

(Eq) $$\frac{X \quad X =_* Y}{Y,}$$

($\supset$E) $$\frac{X \supset Y \quad X}{Y.}$$

Suppose also that

(PW) $$(X \supset (X \supset Y)) \supset (X \supset Y)$$[24]

is provable for arbitrary terms X and Y. Then any term can be proved in the system. To see this, let Y be any term, and let $X \equiv \mathsf{Y}(\lambda x.x \supset (x \supset Y))$, where Y is a fixed-point combinator,[25] so that $X =_* (X \supset (X \supset Y))$. Then we can prove Y as follows:

1.	$(X \supset (X \supset Y)) \supset (X \supset Y)$	By hypothesis
2.	$X \supset (X \supset Y)$	By 1 and Rule (Eq)
3.	X	By 2 and Rule (Eq)
4.	$X \supset Y$	By 2, 3, and Rule ($\supset$E)
5.	Y	By 3, 4, and Rule ($\supset$E)

Before he discovered this paradox, Curry had assumed that the contradiction was caused by his postulates for the universal quantifier, but this paradox made it clear that the problem was the postulates for implication alone.[26]

Curry's next step was to set out a program to find type-free systems based on combinatory logic or λ-calculus that would be consistent. His idea was to define a restricted class of terms that could represent propositions.

[23] [Curry and Feys, 1958, §8S1].

[24] This will be provable if, in addition to the above rules, the natural deduction rule ($\supset$I) for introducing implication (also known as the Dedution Theorem) is either a primitive rule or a derived rule.

[25] See Appendix A.

[26] For further confirmation of this, see [Seldin, 2000].

For this purpose, he noted that there are different possible sets of logical operators that could be taken as primitive.

First, he could use P and Π, as he had done in his early papers. In order to meet Rosser's criterion that the conclusion of every primitive rule should be uniquely determined by the premises, he added a new atomic constant E with the property that $\mathsf{E}X$ is provable for every X. Then he characterized P and Π by the rules

Rule P. $\mathsf{P}XY, X \vdash Y$,

Rule Π. $\Pi X, \mathsf{E}Y \vdash XY$.

A second possible set consisted of Church's Π, which Curry renamed Ξ. Its characteristic rule is

Rule Ξ. $\Xi XY, XU \vdash YU$.

In terms of Ξ, both P and Π can be defined as follows:

$$\begin{aligned} \mathsf{P} &\equiv \lambda xy \,.\, \Xi(\mathsf{K}x)(\mathsf{K}y), \\ \Pi &\equiv \Xi\mathsf{E}. \end{aligned}$$

(Because Curry accepted the combinator K, he did not need a separate atomic term for implication the way Church did.) Conversely, Ξ can be defined in terms of P and Π by

$$\Xi \equiv \lambda xy \,.\, \Pi(\lambda u \,.\, \mathsf{P}(xu)(yu)),$$

or, in a more commonly recognized notation,

$$\Xi \equiv \lambda xy \,.\, (\forall u)(xu \supset yu).$$

A third set of primitives consisted of just F, which is characterized by the following rule:

Rule F. $\mathsf{F}\alpha\beta X, \alpha U \vdash \beta(XU)$.

Of course, F can be defined in terms of P and Π by

$$\mathsf{F} \equiv \lambda xyz \,.\, \Pi(\lambda u \,.\, \mathsf{P}(xu)(y(zu))).$$

or, in the more commonly recognized notation,

$$\mathsf{F} \equiv \lambda xyz \,.\, (\forall u)(xu \supset y(zu)).$$

Using essentially the same definition, F can be defined in terms of Ξ:

$$\mathsf{F} \equiv \lambda xyz \,.\, \Xi x(\lambda u \,.\, y(zu)).$$

But Ξ can also be defined in terms of F. In fact, it can be done in two ways.

$$\Xi \equiv \lambda xy \,.\, \mathsf{F}xy\mathsf{I},$$

and

$$\Xi \equiv \lambda xy \,.\, \mathsf{F}x\mathsf{I}y.$$

Curry then proposed to study three different classes of systems:

1. $\mathcal{F}_1$: systems of *functionality*, based on the logical operator F and Rule F,

2. $\mathcal{F}_2$: systems of *restricted generality* based on Ξ and Rule Ξ, and

3. $\mathcal{F}_3$: systems of *universal generality* based on P and Π and Rules P and Π.

For each kind of system, Curry proposed to find restricted forms of introduction rule(s) corresponding to the characteristic rules read as elimination rules, in the sense which is standard in natural deduction. Thus, for systems $\mathcal{F}_1$, he wanted to find a class of terms called *canonical*[27] for which he could prove the following theorem, where lower case Greek letters represent canonical terms.

THEOREM 1 Stratification Theorem. *If* Γ *is any set of assumptions, and* $\Gamma, \alpha x \vdash \beta X$*, where the variable* x *does not occur free in* Γ*,* α*, or* β*, then* $\Gamma \vdash \mathsf{F}\alpha\beta(\lambda x \,.\, X)$*.*

For systems $\mathcal{F}_2$, Curry wanted to define a class of canonical terms from which he could prove:

THEOREM 2 Deduction Theorem for Ξ. *If* Γ *is any set of assumptions and* $\Gamma, \alpha x \vdash \beta$*, where* x *does not occur free in* Γ *or* α*, then* $\Gamma \vdash \Xi\alpha(\lambda x \,.\, \beta)$*.*

Similarly, for systems $\mathcal{F}_3$, he wanted to define a class of canonical terms from which he could prove the following two theorems:

THEOREM 3 Deduction Theorem for P. *If* Γ *is any set of assumptions, and* $\Gamma, \alpha \vdash \beta$*, then* $\Gamma \vdash \mathsf{P}\alpha\beta$*.*

THEOREM 4 Universal Generalization. *If* Γ *is any set of assumptions, and* $\Gamma \vdash \alpha$*, where* x *does not occur free in* Γ*, then* $\Gamma \vdash \Pi(\lambda x \,.\, \alpha)$*.*

Of course, Curry wanted consistency proofs for all these systems. To obtain them, he proposed to use systems like those of Gentzen and prove results corresponding to the *Cut Elimination Theorem* [Gentzen, 1934].[28]

In his earlier versions of the above systems, the axioms of each system were presented as axiom schemes. But Curry was also interested in replacing these schemes by single axioms. This meant incorporating the definition of canonical terms into the logic using a new atomic constant H to indicate canonicalness. For each kind of system $\mathcal{F}_i$, Curry wanted to define a corresponding finitely axiomatized system $\mathcal{F}_i^*$.

Then the United States entered World War II. For a combination of patriotic and financial reasons, Curry took a leave of absence from Penn State starting in June, 1942, in order to do applied mathematics for the U. S. government. Before going, he wrote up his current research results and published them as [Curry, 1942a; Curry, 1942c].

Just before his leave of absence began, he finally came to understand the role of the combinator S after reading [Rosser, 1942]. Rosser had based his postulates on

[27] Curry actually called them *canonical obs* or *canobs*. This is because he was using the word 'ob' to refer to terms. See Definition 1 of Section §3 below.

[28] See Section §3 below.

his favorite combinators, I and J. But most important for Curry's understanding of S, Rosser had defined abstraction by induction on the structure of the term. When Curry translated Rosser's postulates into the more usual combinators, he finally saw that S could be used in a very elegant definition of bracket abstraction (see Appendix C). Because Curry was about to leave Penn State for his war effort, he was unable to write this up as a paper, so he sent his notes to Rosser.

Actually, the idea of defining abstraction by a direct induction was not really due to Rosser. The definition in [Rosser, 1942] had originally been given by Church for combinations in λ-calculus in [Church, 1935, §3] and in [Church, 1941, Chapter IV]. But it was [Rosser, 1942] that stimulated Curry to see how to use S.

2.5 Curry's War Work and Combinatory Logic

Most of the work Curry did during World War II was unrelated to combinatory logic. The greater part of it was on the fire control problem, which is the mathematics of aiming a projectile at a moving target, and he also did some work with I. J. Schoenberg on splines [Curry and Schoenberg, 1947; Curry and Schoenberg, 1966]. However, at the end of the war and immediately after it, in late 1945 and early 1946, Curry was involved with the ENIAC[29] computer project. This eventually led to the idea of using combinators to combine computer programs [Curry, 1954].

Curry's experience with the ENIAC project also had an after-effect on his return to Penn State in September 1946. He tried to pursuade the administration to get some computing equipment and start what is now called a computer science program. He had no success in this, but persisted until a colleague pointed out to him that if he did succeed, he would be made director of the program without any increase in salary. He then gave up trying to start such a computer science program at Penn State, and went back to logic.

2.6 Curry's Work on Combinatory Logic in the Early Postwar Years

The first thing Curry did about combinatory logic after returning to Penn State was to write to Rosser about the notes he had sent him on basing combinatory logic on K and S and defining abstraction by induction on the structure of the term. He wanted to know whether Rosser wanted to write a joint paper about this. But Rosser had moved on to other things, so Curry then wrote his own paper [Curry, 1949].

About the same time, Curry became aware of [Newman, 1942], and thought that he could use the ideas in it to give a new proof of the Church-Rosser Theorem in an abstract setting. In a sense he succeeded in [Curry, 1952c], since he did give a proof of an abstract version of the theorem, but after the paper appeared it

[29]The ENIAC was one of the first electronic computers, built during World War II for the purpose of calculating the trajectories of projectiles. See [Goldstine, 1946], now available at `http://ftp.arl.mil/ mike/comphist/46eniac-report/index.html`.

was discovered that λ-calculus does not satisfy one of the hypotheses he assumed. (And Newman's proof failed for a similar reason.)

Meanwhile, Curry attended the Tenth International Congress of Philosophy in Amsterdam in the summer of 1948. During that congress, he was approached by one of the editors of the new North-Holland Publishing Company and asked to write a short monograph of at most 100 pages on combinatory logic for their series "Studies in Logic". Curry sent them a philosophical manuscript he had ready, [Curry, 1951]. But, for combinatory logic, he realised that so much unpublished research had accumulated that an adequate treatment would need to be much longer than the editors envisaged.

Nevertheless, the idea of writing a book on this subject appealed to him. He knew that he was not a good expository writer, so he approached Robert Feys, who was on the faculty of the University of Louvain in Belgium and had an interest in combinatory logic, to become a co-author. Feys accepted, and Curry applied for and received a Fulbright fellowship to spend the year 1950–1951 at Louvain. There, work began on [Curry and Feys, 1958].

From the beginning, the plan was for a two volume work, with the first volume dealing with properties of conversion, called *pure combinatory logic*, and the second volume dealing with systems of logic in the ordinary sense based on combinatory logic or λ-calculus, called *illative combinatory logic*. However, as the writing proceeded, it became clear that this plan needed to be modified: in particular, since Curry was continuing to do new research, some of his newer results on the theory of functionality came to be included in the first volume.

Thus, [Curry and Feys, 1958] started with introductory material on formal systems (Chapter 1) and metatheory[30] (Chapter 2), and then had a chapter (Chapter 3) on λ-calculus. Although the λ-calculus was basically due to Church, it was Curry who fixed some notations that became standard in subsequent works on λ-calculus, for example the names for alpha-reduction and beta-reduction. He used the lower-case Greek alphabet for names of binary relations, and (α) and (β) became his designations for Church's Rules I and II from [Church, 1932]. The chapter on λ-calculus was followed by a chapter (Chapter 4) on the Church-Rosser Theorem.[31] This was followed by three chapters on pure combinatory logic: Chapter 5 was on various commonly used composite combinators and the relations between them, Chapter 6 was on what is called the "synthetic" theory of combinators (formal definition of the system, definition of abstraction, and the relation to λ-calculus), and Chapter 7 was on what was called "logistic foundations," which means formulations in terms of provability. Chapter 8 was a general introduction to illative combinatory logic, and Chapters 9–10 were on the theory of functionality.

The new research included in [Curry and Feys, 1958][32] grew out of work on

[30] There called "epitheory" for reasons which will be discussed in § 4 below.

[31] The proof presented in this chapter has since been superceded by a simpler proof; see [Hindley and Seldin, 1986, Appendix 1].

[32] A careful study of Curry's notes from the period and also of the manuscript that he had

the theory of functionality, that is, systems of type $\mathcal{F}_1$. Curry had originally hoped that a theory of functionality in which all terms were canonical would be consistent, and beginning in April 1954, he spent a considerable amount of time trying to prove this. His approach was to show first that all inferences by Rule Eq could be postponed to the end of any deduction and then to show that any deduction in which the only rule was Rule F could be transformed into a standard form. Deductions in which the only rule is Rule F are called *F-deductions*. During the summer of 1954, he noticed if ξX could be deduced by an F-deduction from a set of assumptions, then X must be irreducible and perhaps in some kind of an even more restricted normal form. At that time, the only kind of normal form he had for combinators was weak normal form, which means a combinatory term with no redexes of the form $\mathsf{K}XY$ or $\mathsf{S}XYZ$. Thus, when he first wrote down the *Normal Form Theorem* in his notes, it took the form "*If ξX can be deduced from a [set of assumptions] B, then X_λ has a normal form.*"[33] It took Curry some time to prove the theorem, and the problem he had stating the theorem for combinatory terms motivated him to introduce a new kind of reduction for combinators,[34] After proving his Normal Form Theorem, he returned to the system whose consistency he had been trying to prove. He realized that it was, in fact, inconsistent, and in [Curry, 1955] he published a contradiction to it. He then investigated some more restricted systems, and found some that he could prove consistent [Curry, 1956]. This material appeared in [Curry and Feys, 1958, Chapters 9–10]. Chapter 9 was devoted to the *basic theory of functionality*, in which the types are those definable from a set of atomic types by means of the operation of forming $\mathsf{F}\alpha\beta$ from α and β. Thus, types in the basic theory of functionality are the types defined in Definition 4 in Appendix B below. The axioms were given by two axiom schemes:

$$(\mathsf{FK}) \qquad \mathsf{F}\alpha(\mathsf{F}\beta\alpha)\mathsf{K},$$

$$(\mathsf{FS}) \qquad \mathsf{F}(\mathsf{F}\alpha(\mathsf{F}\beta\gamma))(\mathsf{F}(\mathsf{F}\alpha\beta)(\mathsf{F}\alpha\gamma))\mathsf{S}.$$

For Curry, the types were just a special kind of terms, but since the atomic types were all assumed to be like F in being constants that do not head redexes, there was no point in considering conversions between types. Thus, Curry restricted the conversion rule, Rule Eq, to conversions between subjects only, and he called this Rule Eq′:

$$(\text{Eq}') \qquad \frac{\alpha X \quad X =_* Y}{\alpha Y}$$

For F-deductions, Curry proved the Subject-Construction Theorem (i.e. that de-

received from Feys shows that this work was all done by Curry with no input by Feys.

[33]The normal form theorem for typed λ-calculus had been stated earlier by Alan Turing in an unpublished manuscript (see [Gandy, 1980]), but, of course, Curry did not know of this.

[34]See Strong Reduction below.

ductions always follow the construction of the subjects),[35] the Subject-Reduction Theorem (see Theorem 13 in Appendix B below), and the Stratification Theorem (see Theorem 1 in Section 2.3 above). He also proved that in a deduction with Rule (Eq′), all inferences by Rule (Eq′) can be postponed until the end and combined into one. In addition, he defined a version of basic functionality for λ-calculus, in which the axiom schemes (FK) and (FS) are replaced by a rule which he called an axiom scheme:

(Fλ) $$\frac{\begin{array}{c}[\alpha_1 x_1, \alpha_2 x_2, \ldots, \alpha_m x_m] \\ \beta X\end{array}}{\mathsf{F}\alpha_1(\ldots(\mathsf{F}\alpha_m\beta)\ldots)(\lambda x_1 \ldots x_m \,.\, X)}$$

Condition: X is a combination of the variables $x_1, x_2, \ldots, x_m$ only, and these variables do not occur free in any undischarged assumption.

Finally, he noted that if the subjects are removed from all steps of a deduction and if each occurrence of F is changed to an occurrence of P, the result is a valid deduction in the implication fragment of intuitionistic propositional calculus.[36] He concluded the chapter with natural deduction and Gentzen-style L-formulations, and he used the latter to give the first published proof of the Normal Form Theorem; see Theorem 14 of Appendix B below).

In [Curry and Feys, 1958, Chapter 10], Curry presented the results of both [Curry, 1955] and [Curry, 1956]. As mentioned before, these included a proof of consistency for a restricted system.

The problem that Curry originally had stating the Normal Form Theorem let him to introduce another innovation in [Curry and Feys, 1958]: *strong reduction* in Chapter 6. In pure combinatory logic, the natural reduction, weak reduction, lacks a property that both $\lambda\beta$- and $\lambda\beta\eta$-reduction satisfy, namely the property

(ξ) $$X \rhd Y \to [x]X \rhd [x]Y.$$

Curry studied the relation $\succ$ defined by adding (ξ) to the definition of weak reduction and inserting into the definition of bracket abstraction (see Appendix C) the clause

(c) if x does not occur free in U, then $[x](Ux) \equiv U$.

He called this relation *strong reduction*. He showed that it corresponded to $\lambda\beta\eta$-reduction, provided the combinator I was assumed to be an atom instead of being

[35] This theorem is based on the way he worked out examples in his attempt to prove the consistency of the "full free" theory of functionality. He later formally defined this as what is now known as a "typing algorithm."

[36] This observation, in extended form, is now well known as the formulas-as-types or propositions-as-types or Curry-Howard isomorphism. See [Howard, 1980].

defined as SKK.[37]

Six years after Curry and Feys began their work, the writing was finished, and [Curry and Feys, 1958] appeared two years later.

Meanwhile, Curry had his first graduate student, Edward J. Cogan. Cogan wrote a dissertation on set theory within illative combinatory logic, [Cogan, 1955]. The idea was to use the ideas about set theory of [Gödel, 1940; von Neumann, 1925; von Neumann, 1928] within the framework developed by Curry.[38] Cogan chose to represent the set theory of [Gödel, 1940] in a system of type $\mathcal{F}_2^*$ based on Ξ, with F defined in terms of Ξ. Unfortunately, his system turned out to be inconsistent [Titgemeyer, 1961]. The problem came from an axiom that Curry had suggested to Cogan.

2.7 Curry's Later Work on Combinatory Logic

After the publication of [Curry and Feys, 1958], Curry himself, in preparation for writing the second volume of [Curry and Feys, 1958] turned his attention to systems of type $\mathcal{F}_2$, systems of restricted generality. In his [1960], Curry proved a version of the Deduction Theorem for Ξ, namely: *if* $\xi_1, \xi_2, \ldots, \xi_m, \eta$ *are all canonical, and if*

$$\Gamma, \xi_1 x_1, \xi_2 x_1 x_2, \ldots \xi_m x_1 x_2 \ldots x_m \vdash \eta,$$

where the variables $x_1, x_2 \ldots, x_m$ *do not occur free in* Γ*, then*

$$\Gamma \vdash \Xi\xi_1(\lambda x_1 \,.\, \Xi(\xi_2 x_1)(\ldots(\lambda x_{m-1} \,.\, \Xi(\xi_m x_1 x_2 \ldots x_{m-1})(\lambda x_m \,.\, \eta))\ldots)).$$

Canonical terms were defined in a such a way that the system was not really stronger than first order logic. Curry was unable to obtain this result by iterating Theorem 2 above because he had failed to assume an axiom generating rule. This problem arises in ordinary predicate calculus, where one needs in addition to modus ponens either a rule of universal generalization or else a specification that the universal closure of any axiom is an axiom. In his original work, Curry left out this specification (which I called above the axiom generating rule), although in the paper he assumed it in the multiple form.

In his [1961], Curry gave a proof of a sort of cut-elimination theorem for a system related to a system of type $\mathcal{F}_2$ the way a Gentzen L-system is related to a natural deduction system.

Meanwhile, Curry had three more Ph.D. students. Kenneth Loewen [1962] wrote on the Church-Rosser Theorem and the Standardization Theorem[39] for strong reduction, and Bruce Lercher [1963] wrote on the relationship between

[37] The definition of a relation corresponding to $\lambda\beta$-reduction was not attempted. This involved some technical complications and was not done until [Mezghiche, 1984; Mezghiche, 1989; Mezghiche, 1997].

[38] In a sense this set-theory already contained a combinator concept in some of its axioms, [Curry and Feys, 1958, p. 10], [Cardone and Hindley, 2008].)

[39] The Standardization Theorem says roughly speaking that any reduction can be carried out by reducing redexes from left to right.

strong reduction and the representation of recursive functions and functionals of higher type. Meanwhile, Luis E. Sanchis [1963] formulated Church's type theory within combinatory logic using the theory of functionality for the types, and proved the consistency of that system.

On April 16, 1961, Robert Feys died. This left Curry alone to work on the second volume of [Curry and Feys, 1958]. He realized that he needed collaborators and also realized that he needed to make a further study of the foundations of logic, and in particular, of Gentzen-style[40] proof theory before writing the second volume. The result of this study was [Curry, 1963]. Thus, he did not get back to work on the second volume until 1964. That September, J. Roger Hindley arrived at Penn State to do postdoctoral work with Curry. Hindley had just completed his dissertation [1964] on the Church-Rosser Theorem, and in 1965 Curry invited him to join the project. Also in September, 1964, I began my graduate studies with Curry. When Curry retired from Penn State in 1966 and took up a position at the University of Amsterdam, I accompanied him. After I finished my dissertation [1968], Curry invited me to join the project as well, and this is how the second volume became [Curry *et al.*, 1972], and some of the material of [Curry *et al.*, 1972, Chapters 14–16] came from [Seldin, 1968, Chapters 3–5]. When Curry arrived in Amsterdam, another graduate student, Martin W. Bunder began studying with him. Curry ran a seminar at which Martin Bunder and I presented our results, and although Martin did not become a co-author of [Curry *et al.*, 1972], his work did contribute to some of the later chapters.

The first chapter of [Curry *et al.*, 1972], Chapter 11,[41] is devoted to updates to the material on pure combinatory logic in the first volume. Curry's original intention was to begin with a chapter on philosophical questions, but that proved too difficult to write, and so only two subsections of the first section of the chapter are devoted to that. In addition to the updates to the material on pure combinatory logic, the chapter contains a proposal due to Curry for a kind of system, which he called a *C-system*, which would treat combinatory logic and λ-calculus simultaneously. The idea was that since the various illative systems can be defined in both λ-calculus and combinatory logic, there should be a kind of system which could be instantiated to either. This idea was applied throughout the book, although in several later places it was still necessary to treat combinatory logic and λ-calculus separately.

Chapter 12 is an introduction to illative systems. Originally, illative combinatory logic meant combinatory logic with operators for the logical connectives and quantifiers, but by the mid-1960s, Curry had decided that any applied combinatory logic or λ-calculus should be considered an illative system.

Chapter 13 is entirely on the representation of natural numbers and recursive functions in combinatory logic. To allow for the various known ways of coding numerals as combinatory or lambda terms, Curry introduced two new atoms **0**

[40] See § 3 below.

[41] The chapter numbering continues from [Curry and Feys, 1958].

and $\boldsymbol{\sigma}$, and postuated what he called an *iterator* Z with the property that

$$\mathsf{Z}\,\underbrace{(\boldsymbol{\sigma}(\boldsymbol{\sigma}(\ldots(\boldsymbol{\sigma}}_{n}\,\mathbf{0})\ldots)) \rhd \lambda xy\,.\,\underbrace{x(x(\ldots(x}_{n}\,y)\ldots)).$$

This system could be instantiated to any of the numerical codings by substituting appropriate terms for $\mathbf{0}$, $\boldsymbol{\sigma}$, and and constructing an appropriate pure combinator to serve as Z. For the coding used by Church, $\mathbf{0}$ and $\boldsymbol{\sigma}$ would be replaced by Church's definitions of those terms and Z would be replaced by I. The chapter included a proof that all recursive functions can be represented. It also discussed the use of terms with types, or functional characters, and ended with a discussion of Gödel's functionals of finite type. Included in the chapter was an undecidability result of Curry's for $\lambda\mathsf{K}$-calculus and equivalent combinatory systems, a result that was stronger than Church's result of his [1936c]. Curry [1969b] proved that the only *conversion-invariant*[42] recursive sets of terms were the empty set and the set of all terms. Since the set of terms which convert to a given term and the set of terms with a normal form were conversion-invariant, this implied Church's two results, but Curry's proof required the combinator K.[43]

Chapter 14 is devoted to the theory of functionality. Curry had, by this time, split Rule (Eq) into separate rules for subject and predicate:

(Eqs) $$\frac{\alpha X \quad X =_* Y}{\alpha Y}$$

(Eqp) $$\frac{\alpha X \quad \alpha =_* \beta}{\beta X}$$

Here, (Eqs) was the rule (Eq′) of [Curry and Feys, 1958, Chapter 9], which is the only one appropriate for basic functionality, but for stronger systems of functionality (Eqp) would also be appropriate. Curry called a system using these rules a *separated system*, and it is with separated systems that functionality finally became what is now recognized as a type theory, in which Curry's type assignment statements αX could be rewritten $X : \alpha$. Chapter 14 also included a number of technical improvements to the systems of [Curry and Feys, 1958, Chapters 9–10]. One of the improvements is the proof that every term with a type has a *principal type scheme* or *principal functional character*. This result was proved by two different methods in [Curry, 1969a] and [Hindley, 1969].

Chapter 15 is devoted to systems of restricted generality, that is systems of type $\mathcal{F}_2$, based on Ξ. In it, a form of the Deduction Theorem for Ξ like Theorem 2 above that could be iterated was proved, and with this and a cut elimination theorem

[42] A set of terms is conversion invariant if, whenever X is in the set and $X =_* Y$, then Y is in the set. Curry called these sets *equation-invariant*.

[43] A very similar undecidability theorem had been proved by D. Scott in lecture notes in 1963, which Curry did not see until his own proof was substantially written. See [Curry, 1969b, p.10] or [Curry *et al.*, 1972, p.251, footnote 7].

for a Gentzen-style L-system, the consistency was proved of a system whose rules for canonicalness are essentially of the strength of first-order logic. This settled the theory of $\mathcal{F}_2$ systems. However, the $\mathcal{F}_2^*$ systems were definitely not settled. Recall that the $\mathcal{F}_2^*$ systems are those in which canonicalness is defined using the theory of functionality (which can be embedded in these systems) in conjunction with a term H representing the predicate of being a proposition. A system of this kind, called $\mathcal{F}_{21}^*$, was proposed in [Seldin, 1968, Chapter 4], using an idea of Martin Bunder to restrict the deduction theorem by placing a restriction only on the antecedent: i.e., to take for the deduction theorem

$$\Gamma, Xx \vdash Yx \;\Rightarrow\; \Gamma, \mathsf{L}X \vdash \Xi XY,$$

where x does not occur free in Γ, X, or Y, and L is FEH, the category of one-place predicates.[44] In [Curry *et al.*, 1972, §15C1] a different system also called $\mathcal{F}_{21}^*$ is defined. Both of these systems were defined to avoid some contradictions found by Martin Bunder in his [1969]. However, as indicated below, neither turned out to be consistent.

Chapter 16 of [Curry *et al.*, 1972] was on systems of universal generality, that is systems of type $\mathcal{F}_3$, based on Π and P. The basic conclusion of this chapter is that systems of universal generality are essentially equivalent to systems of restricted generality as long as E is present in the latter. This contradicts Curry's original idea that $\mathcal{F}_3$ systems might be stronger than $\mathcal{F}_2$ systems. Curry also originally assumed that both $\mathcal{F}_2$ and $\mathcal{F}_3$ systems were stronger than $\mathcal{F}_1$ systems, but if $\mathcal{F}_1$ systems are taken without the separated equality rules and Ξ is defined in terms of F in either of the ways indicated above, the result is equivalent to the corresponding $\mathcal{F}_2$ system. Thus, there are equivalent versions of all three kinds of systems.

Chapter 17 was on type theory based on combinatory logic and the theory of functionality. The chapter extended some of the results of [Sanchis, 1963; Sanchis, 1964]. One of the extensions was to a type theory with transfinite types, as in [Andrews, 1965].

The writing of [Curry *et al.*, 1972] was finished in May of 1970, and after that Curry retired from the University of Amsterdam and returned to State College, Pennsylvania. Except for the year 1971–1972, he spent the rest of his life there. But before he left Amsterdam, his last Ph.D. student, Martin Bunder, finished his dissertation [Bunder, 1969]. Bunder's thesis showed how to interpret both ZF and NBG set theories in a system of restricted generality. His system was similar to the two systems $\mathcal{F}_{21}^*$ of Seldin and Curry, but he defined L to be FAH, where A is a category of individuals and it is not the case that $\vdash \mathsf{A}X$ for every term X.

In his [1969, Chapter 2], Bunder gave a number of contradictions that could be derived in various illative systems as a motivation for the limitations of his own system (for which there was no consistency proof). Bunder later proved Seldin's system $\mathcal{F}_{21}^*$ inconsistent in his [1974]. This proof of inconsistency does not apply to the system $\mathcal{F}_{21}^*$ of [Curry *et al.*, 1972], but in his [Bunder, 1976] Bunder proved

[44] All previous proposals for this kind of system had restricted the deduction theorem by means of restrictions on both the antecedent and the consequent.

this system inconsistent as well. Finally, in their [1978], Bunder and Meyer proved inconsistent any system of illative combinatory logic which includes the following:

1. Rule (Eq),
2. Modus ponens: $\mathsf{P}XY, X \vdash Y$,
3. A deduction theorem as follows: $\Gamma, X \vdash Y \Rightarrow \Gamma, \mathsf{H}X, \mathsf{H}Y \vdash \mathsf{P}XY$.
4. $\mathsf{H}X, \mathsf{H}Y \vdash \mathsf{H}(\mathsf{P}XY)$,
5. $X \vdash \mathsf{H}X$, and
6. $\vdash \underbrace{\mathsf{H}(\mathsf{H}(\ldots(\mathsf{H}}_{n} X)\ldots))$ for a fixed n and every X.

The case $n = 2$ of item 6, namely $\vdash \mathsf{H}(\mathsf{H}X)$ for every X, was of interest because it is equivalent to $\vdash \mathsf{LH}$, which is necessary to derive any reasonable set of axioms from the deduction theorem. Finding a set of axioms equivalent to the deduction theorem had been one of Curry's major aims ever since he formulated his new program in the 1930s, and this result proved that impossible.

But even before Bunder published his contradictions, Curry became suspicious of $\vdash \mathsf{H}(\mathsf{H}X)$ for every X. His reason was that if $\vdash \mathsf{H}(\mathsf{H}X)$ holds for every term X, then it must hold for $X \equiv \mathsf{YH}$, where Y is a fixed point combinator, and from this follows $\vdash \mathsf{YH}$, which he thought was counterintuitive. For this reason, Curry gave up on trying to find a system with axioms equivalent to the deduction theorem, and, in his [1973], defined a system he called $\mathcal{F}_{22}$, which was the system $\mathcal{F}_{21}^*$ of [Curry *et al.*, 1972, §15C] without the postulate $\vdash \mathsf{H}(\mathsf{H}X)$, and proved it consistent in the weak sense that every theorem is canonical. Seldin obtained stronger consistency results for this system in his [1975b; 1977] by proving cut elimination, thus showing that there are canonical terms which are not provable. This system is defined as follows:

Definition 1. *The* system $\mathcal{F}_{22}$ *is based on an applied* $\lambda\beta\eta$*-calculus or equivalent combinatory logic with atomic constants (that do not head redexes)* Ξ, E, H, *and perhaps additional canonical atoms (i.e., atomic predicate functions) of degree (number of arguments)* n. *The term* F *is defined by*

$$\mathsf{F} \equiv \lambda xyz\,.\,\Xi x(\lambda u\,.\,y(zu)).$$

The axioms are $\mathsf{E}X$ *for all terms* X *and also* FEHE. *The rules are as follows:*

(Eq) $$\frac{X \quad X =_* Y}{Y}$$

(ΞE*)* $$\frac{\Xi XY \quad XU}{YU}$$

(ΞI)	$\dfrac{\begin{array}{c}[Xx]\\ Yx \quad \mathsf{FEH}X\end{array}}{\Xi XY}$	Condition: *x does not occur free in X, Y, or in any undischarged assumption.*
(H)	$\dfrac{X}{\mathsf{H}X}$	
(Hθ)	$\dfrac{\mathsf{E}X_1 \quad \mathsf{E}X_2 \quad \dots \quad \mathsf{E}X_n}{\mathsf{H}(\theta X_1 X_2 \dots X_n)}$	Condition: *θ is a canonical atom of degree n.*
(HΞ)	$\dfrac{\mathsf{FEH}X \quad \mathsf{FEH}Y}{\mathsf{H}(\Xi XY)}$	

This was the last illative system about which Curry published any results. However, he did publish several more papers on pure combinatory logic. In his [1975b], Curry proved for a generalized reduction of the kind used in [Curry *et al.*, 1972, §13A] that if a term has a normal form, then it reduces to that normal form.[45] In his [1975a], he showed how to represent Markov algorithms in combinatory logic. His [1979] is on the representation of terms in normal form in the $\lambda\beta$-calculus. And his [1980], Curry's last published paper, is a brief survey of combinatory logic followed by some responses to some philosophical criticisms of it.

Martin Bunder has done more than anybody else to carry on Curry's tradition of illative combinatory logic. His most recent results are joint work with Wil Dekkers and Henk Barendregt [Barendregt *et al.*, 1993; Dekkers *et al.*, 1998a; Dekkers *et al.*, 1998b], and include the consistency proof of a stronger system than any system previously proved consistent.

3 CURRY'S WORK ON PROOF THEORY

When Curry first read [Gentzen, 1934], he immediately realized its importance. When he formulated his new program after learning of the paradox of [Kleene and Rosser, 1935], he realized that Gentzen's techniques might be useful in obtaining the consistency proofs that his new program called for.

As a result, he started writing a paper, of which [Curry, 1937] is the abstract, on the subject. However, this paper was never finished,[46] and all he published on this subject before World War II was [Curry, 1939a].

After the war, Curry was invited to give a series of lectures at Notre Dame University April 12–15, 1948, and he chose as his subject an exposition of Gentzen's work. Instead of revising his earlier paper, he started over, and the result was ultimately published as [Curry, 1950].

[45] The generalized reduction was supposed to be that for what Curry called a C-system.

[46] Only the first few sections were ever typed; the rest exists only as handwritten notes.

In this work, Curry started from the point of view of what we might call a *Curry semantics*. Since Curry has a reputation for not having paid attention to semantical considerations, some may find this notion surprising, but Curry did, in fact, have definite ideas about the meaning of the logical connectives and quantifiers, and if they have been misunderstood it is because they are not based as model theory is on set theory. Curry's idea behind his exposition was that he was formalizing a part of the metatheory of an elementary formal system, and he justified the properties of the logical connectives and quantifiers on that basis.

Curry defined a *formal system* as follows:

Definition 2. *A* formal system[47] *is a set*[48] *E of statements, called* elementary statements*, together with a subset* $T \subseteq E$ *of* theorems*, where the two sets satisfy the following conditions:*

1. *The set T is* inductive*, that is, it is generated by a* definite[49] *set of* axioms *by a set of* rules *in such a way that there is an effective procedure which can decide, given a sequence of elementary statements, whether or not that sequence constitutes a valid proof.*

2. *The set E is formed from a definite set O of* formal objects *and a definite set of* elementary predicates*, each with a fixed number of arguments. A statement is in E if and only if it asserts that an elementary predicate of n arguments applies to an n-tuple of formal objects.*

Formal systems are classified according to the way the set O of formal objects is formed.

1. *If the formal objects are defined from primitive objects, called* atoms*, and* primitive operations *in such a way that each formal object has a unique construction, then the system is called an* ob-system*, and the formal objects are called* obs.[50]

2. *If the formal objects are words (strings of characters) on some alphabet,*[51]

[47] This is a late form of Curry's definition, taken from [Curry, 1963]. However, the ideas behind this definition date from the beginning of his career. For an exposition of Curry's notion of formal system, see [Seldin, 1975a].

[48] Here the word "set" is taken to mean a conceptual class, and its use does not presuppose any formal theory of sets.

[49] Curry called a set *definite* if there is an effective procedure for deciding whether or not a given object of a suitable universe is a member of the set.

[50] This is a word that Curry coined for this purpose about 1950. In his earliest papers, he called them *entities*, but when a philosopher told him that the word "entity" implied a meaning that he did not intend, he switched to the word *term*. However, the word "term" has another use in connection with a system of logic with quantifiers, and Curry felt in need of another word. In this paper, I have used the word "term" rather than "ob" in discussing λ-calculus and combinatory logic because the former is more common.

[51] In general, such words do not have unique constructions. For example, the word 'abc' can be constructed by adding 'a' to the front of 'bc' or by adding 'c' to the end of 'ab'.

then the system is called a syntactical system.[52]

An elementary formal system *is a formal system in which all the rules have the form*

$$\frac{E_1 \quad E_2 \quad \dots \quad E_n}{E_0,}$$

where $E_0, E_1, E_2, \dots, E_n$ *are schemes for elementary statements; i.e., are expressions that become elementary formal systems when formal objects are substituted for (meta-)variables.*

Ordinary formal logical calculi, such as propositional calculus and first-order predicate calculus, are both syntactical systems and ob systems, since the formal objects are defined to be words on an alphabet, but they are also *well-formed formulas*, and well-formed formulas defined in the usual way each have a unique construction from the atomic formulas by means of the connectives and quantifiers.

For Curry, the elementary statements of logical calculi are defined by applying the predicate 'is provable' to well-formed formulas. Curry indicated this by writing the sign '$\vdash$' in front of each well-formed formula. Thus, whereas it is usual in logic to take a proof as a sequence of well-formed formulas $A_1, A_2, \dots, A_n$, Curry took a proof as a sequence of elementary statements of the form $\vdash A_1, \vdash A_2, \dots \vdash A_n$, and he wrote

$$P_1, P_2, \dots, P_m \ \vdash \ Q$$

as an abbreviation for the statement: if $\vdash P_1, \vdash P_2, \dots, \vdash P_m$ are added to the system as new axioms, then $\vdash Q$ is provable.[53]

Among logical calculi, formal systems of what is usually called a *Hilbert-style formulation*, in which the only rules are modus ponens and universal generalization or modus ponens alone, are elementary formal systems. Natural deduction systems, in which there are rules which discharge assumptions, are not elementary formal systems.

Now Curry had originally defined combinatory logic in his [1929] as an ob system in which the combinatory terms are the obs and the elementary predicate is a predicate of two arguments which he denoted by '=', so that the elementary statements all had the form $X = Y$,[54] where X and Y are combinatory terms. However, starting in his [1930], Curry defined combinatory logic to be what he later called an *assertional system*, a system in which there is only one predicate, namely $\vdash$. Thus, the elementary statements all asserted the provability of combinatory terms, and $X = Y$ became an abbreviation for $\mathsf{Q}XY$.

Curry considered the logical calculi he introduced in his [1950] to be formalizations of a part of the metatheory[55] of an elementary formal system. He thus defined the following metatheoretic connectives:

[52]Originally, Curry considered only ob systems as formal systems. However, starting in his [1963], Curry included syntactical systems as formal systems.

[53]If $m = 0$, so that there are no new axioms added, this just asserts that $\vdash Q$ is provable in the existing system.

[54]There, $X = Y$ had the meaning now denoted by $X =_* Y$.

[55]Kleene, in a review of [Curry, 1941d] published in the *Journal of Symbolic Logic*, 6:10–102,

& : $E_1 \& E_2$ means that there is a proof of E_1 and a proof of E_2,

or: E_1 or E_2 means that there is a proof of E_1 or a proof of E_2 and there is an effectively computable process for deciding which,

$\rightarrow$: $E_1 \rightarrow E_2$ means that there is an effective process for converting any proof of E_1 into a proof of E_2, and is regarded as vacuously true if it can be constructively proved that there is no proof of E_1,

$\rightleftarrows$: $E_1 \rightleftarrows E_2$ means $(E_1 \rightarrow E_2)$ & $(E_2 \rightarrow E_1)$,

$\Rightarrow$: $E_1 \Rightarrow E_2$ means that if E_1 is adjoined to the system as an axiom, then E_2 is a theorem, and

$\Leftrightarrow$: $E_1 \Leftrightarrow E_2$ means $(E_1 \Rightarrow E_2)$ & $(E_2 \Rightarrow E_1)$.

(Clearly, these connectives can connect all statements formed by starting with elementary statements and using these connectives. And it is in terms of their meaning that Curry justified the rules of the connectives $\wedge$, $\vee$, and $\supset$.)[56]

Thus, his [1950] began with a chapter on formal systems, and throughout the rest of the book the logical systems were designed to formalize the metatheory of an elementary formal system $\mathfrak{S}$. In the second chapter, he treated the propositional calculus without negation, so the connectives were $\wedge$, $\vee$, and $\supset$. The first logical systems introduced were the Gentzen-style L-systems, LA($\mathfrak{S}$) and LC($\mathfrak{S}$). The first, LA($\mathfrak{S}$), was a singular system, one with only one formula on the right:

$$\Gamma \Vdash A,$$

where Γ is a sequence of formulas and A is a formula. The second, LC($\mathfrak{S}$), allowed sequences with more than one formula on the right:

$$\Gamma \Vdash \Delta,$$

where Γ and Δ are sequences of formulas. Both kinds of systems had the usual axiom $A \Vdash A$, and both also had as axioms $\Vdash E$, where E was an axiom of $\mathfrak{S}$. Both had the usual structural rules of permutation, weakening, and thinning, although LA had them only on the left while LC had them on the right as well, and both had the usual rules for $\wedge$, $\vee$, and $\supset$ on both the left and right. Both also had the

1941, objected to Curry's use of the prefix "meta-", claiming that this prefix should only be used in connection with a distinction of levels of language. In 1951, Curry responded to this criticism by deciding to use the prefix "meta-" only in the sense Kleene had suggested and to use the prefix "epi-" instead. See the preface of [Curry, 1951]. I think that most logicians now use the prefix "meta-" in the sense in which Curry had used it prior to 1951, and so I will continue to use that prefix instead of "epi-" in this article.

[56] This reading of the connectives is close to the Brouwer-Heyting-Kolmogorov interpretation of them. See [Troelstra and van Dalen, 1988, p. 31].

following rule:

$$\vdash * \qquad \frac{\Gamma \Vdash E_1, \Theta \quad \Gamma \Vdash E_2, \Theta \quad \dots \quad \Gamma \Vdash E_n, \Theta}{\Gamma \Vdash E_0, \Theta} \qquad \begin{array}{l}\textit{Condition:}\\ E_1, E_2, \dots, E_n \vdash E_0\\ \text{in } \mathfrak{S}.\end{array}$$

(In LA($\mathfrak{S}$), Θ was the void sequence.[57]) Curry did not postulate the rule Cut as part of either system, but proved as a metatheorem that if the premises of Cut were derivable, then so was the conclusion. He called this result the Elimination Theorem.

Curry then turned to natural deduction formulations. Since he wanted to reserve the letter 'N' for negation, he used the second consonant of the word "natural" and called the systems TA($\mathfrak{S}$) and TC($\mathfrak{S}$). The only axioms in either system were the axioms of $\mathfrak{S}$. The rules of TA($\mathfrak{S}$) were the standard introduction and elimination rules for $\wedge$, $\vee$, and $\supset$. TC($\mathfrak{S}$) was obtained from TA($\mathfrak{S}$) by adjoining one rule:

$$\supset\text{K} \qquad \frac{\begin{array}{c}[A \supset B]\\ A\end{array}}{A.}$$

Curry also considered Hilbert-style formulations of both systems, HA and HC. These were formulations in which the only rule was modus ponens, or $\supset E$, and the other rules of TA and TC were replaced by axiom schemes.

In the next chapter, Curry considered adding first-order quantification to obtain LA*($\mathfrak{S}$) and LC*($\mathfrak{S}$). The L-rules on the left and right for the quantifiers were standard. The T-rules and H-axioms were also standard.

In the next chapter, Curry took up negation. He defined two kinds of negation:

1. *Absurdity.* An elementary statement is *absurd* if, when it is adjoined to the system as an axiom, all elementary statements become provable.

2. *Refutability.* An elementary statement is *refutable* if, when it is adjoined to the system as an axiom, any one of a given set of *counteraxioms* becomes provable.

These two kinds of negation led to different systems:

1. *Minimal Logic*, LM, which was LA plus refutability.

2. *Intuitionistic Logic*, which was LA plus absurdity.

3. *Strict Refutability*, LD, which was LA plus excluded middle.

4. *Classical Logic*, LK, which was LC plus absurdity, or alternatively LJ plus excluded middle.

[57] If $n = 0$, this means that if $\vdash E_0$ is an axiom of $\mathfrak{S}$, then $\Vdash E_0$ is an axiom of the L-system.

These were the systems of propositional logic; there were corresponding systems with quantifiers.

For technical reasons, Curry was unable to use the definition

$$\neg A \equiv A \supset \bot,$$

where $\bot$[58] stands for a fixed false proposition, in the L-formulation, so in this formulation, he took $\neg$ as a primitive connective. The L-rules for negation on the left and right, $*\neg$ and $\neg*$, were the traditional ones. For refutability, he also postulated

$\bot*$ $\qquad \dfrac{\Gamma \Vdash \bot_i, \Theta}{\Gamma \Vdash \Theta}$ $\qquad$ *Condition:* $\bot_i$ is a counteraxiom.

For absurdity, he postulated instead[59]

$\bot$J $\qquad \dfrac{\Gamma \Vdash \Theta}{\Gamma \Vdash A, \Theta.}$

To get the law of the excluded middle, he postulated

$\neg$X $\qquad \dfrac{\Gamma, \neg A \Vdash A, \Theta}{\Gamma \Vdash A, \Theta.}$

In these rules, Θ was void in a singular system. Note that this meant that in systems with negation, it was possible in either singular or multiple systems to have a void right-hand side.

Then he defined the following propositional systems:

$$\text{LM} = \text{LA} + *\neg + \neg* + \bot*,$$

$$\text{LJ} = \text{LM} + \bot\text{J},^{60}$$

$$\text{LD} = \text{LM} + \neg\text{X}, \text{ and}$$

$$\text{LK} = \text{LC} + *\neg + \neg*.$$

Note that LM, LJ, and LD were all singular. Curry was dealing with quantified versions of all these sytems, and he proved the Elimination Theorem for all of them.

For the natural deduction systems, he used the definition $\neg A \equiv A \supset \bot$, where $\bot$ is a new atomic formula added to the system.[61] The standard introduction and

[58] Curry wrote 'F' instead of '$\bot$'.

[59] There were no counteraxioms in any system in which $\bot$J was postulated, so rule $\bot*$ was never postulated in such systems.

[60] But see the preceding footnote.

[61] In some systems, $\bot$ can be identified with a formula that is already in the system.

elimination rules for ¬, ¬I and ¬E, then follow from the definition of $\neg A$. The other T-rules used in connection with negation were the following:

⊥I	$\dfrac{\bot_i}{\bot}$	*Condition:* $\bot_i$ is a counteraxiom.
¬J	$\dfrac{\bot}{A.}$	
¬D	$\dfrac{\begin{array}{c}[\neg A]\\ A\end{array}}{A.}$	
¬K	$\dfrac{\neg\neg A}{A.}$	

Note that the rule ¬K is equivalent to

$$\dfrac{\begin{array}{c}[\neg A]\\ \bot\end{array}}{A,}$$

which is the form used by Prawitz in his [1965]. The T-systems for negation are then defined as follows:

TM = TA + ¬E + ¬I + ⊥I,

TJ = TM + ¬J,[62]

TD = TM + ¬D, and

TK = TM + ¬J + ¬D = TM + ¬K.

Curry also dealt with Hilbert-style systems.

The last chapter was devoted to modal logic, with operators for necessity and possibility. Curry was unable to prove the Elimination Theorem for the modal systems.

In the preface, Curry discussed the algebraic approach to logic, saying that although he thought it was important, he had to leave it out of the lectures on which [Curry, 1950] was based, for lack of space. In fact, the book did not include answers to some questions left open in the lectures which he had settled after he delivered the lectures, since he wanted the book to represent the lectures themselves. But he did write up this work for publication:

[62]There are no counteraxioms for TJ or for any system in which ¬J is postulated, so ⊥I is vacuous for TJ and TK.

1. In his [1952f], Curry proved some results about LD. He also suggested a singular version of LC and LK, namely LC_1 and LK_1. The former was to be defined by adding to LA the rule

$$\supset\mathrm{C} \qquad \frac{\Gamma, A \supset B \Vdash A}{\Gamma \Vdash A.}$$

   For LK_1, the rules $\bot\mathrm{J}$ and $\neg\mathrm{X}$ were postulated.

2. In his [1952d], Curry showed how to use the definition $\neg A \equiv A \supset \bot$ with the L-systems. When this was done there were no void right-hand sides, and rules $\bot*$, $\bot\mathrm{J}$, and $\neg\mathrm{X}$ were replaced by, respectively,

$$\bot* \qquad \frac{\Gamma \Vdash \bot_i, \Theta}{\Gamma \Vdash \bot, \Theta} \qquad \textit{Condition:} \ \bot_i \text{ is a counteraxiom.}$$

$$\bot\mathrm{J} \qquad \frac{\Gamma \Vdash \bot, \Theta}{\Gamma \Vdash A, \Theta.}$$

$$\bot\mathrm{X} \qquad \frac{\Gamma, A \supset \bot \Vdash A}{\Gamma \Vdash A.}$$

   where, as usual, Θ was void in a singular system.

3. In his [1952e], Curry showed that under certain conditions, the order in which rules are applied can be reversed in classical L-systems.

4. In his [1952a], Curry proved the Elimination Theorem for systems with a necessity operator.

In his review of the first two of the above papers, Bernays [1953] suggested adding $\supset$C to LM without adding $\bot$J. This would be a system of classical refutability, which Curry later called LE. Some properties of this system were studied based on its H-formulation in [Kanger, 1955], and an extensive study starting from the L-formulation was made in [Kripke, 1958]. Although Curry was busy writing [Curry and Feys, 1958] during most of this period and was therefore unable to pursue his work on general proof theory, he kept up with these results.

Once [Curry and Feys, 1958] was finished, Curry decided that since he expected to make extensive use of Gentzen techniques to obtain consistency proofs in the projected second volume, he should put his work on those techniques in better order than it was. This work led to [Curry, 1963]. In this work, after a short introductory chapter, there were chapters on formal systems (Chapter 2) and epitheory (his word for metatheory, Chapter 3), which were largely rewritten from [Curry and Feys, 1958]. Then, in Chapter 4, he included the material on algebraic logic that he had lacked room for in his [1950]; he had been able to give a course on

this subject during his year at Louvain, and the course notes had been published as [Curry, 1952b], and [Curry, 1963, Chapter 4] was expanded from this.

The real work on logical calculi begins in Chapter 5. Unlike [Curry, 1950], Curry started here with Hilbert-style and natural deduction systems. Chapter 5 was on propositional logic without negation, and included LA and LC. But it also included LC_1, and, to this, Curry added a multiple version of LA, LA_m, which was like LC except that the rule for implication on the right must be singular. The chapter closed with a system for tableaux, inspired by Beth's proof tableaux.

Chapter 6 is on negation, and here Curry added the E-systems, HE, TE, and LE to the systems he had previously. TE was defined by adding the introduction and elimination rules for $\neg$ to TC, and this meant that TK could be obtained by adding $\neg$J to TE. LE was defined from LC the way LM was defined from LA, by adding the rules for negation on the left and right and also adding $\bot*$. For LD, Curry replaced rule $\neg$X by a new one:

$$\frac{\Gamma, \neg A \Vdash \Delta \quad \Gamma, A \Vdash \Delta}{\Gamma \Vdash \Delta.}$$

He also had both singular and multiple formulations for all systems. Another innovation was three formulations of negation: 1) the F-formulation, where $\bot$ was postulated and $\neg A$ was defined to be $A \supset \bot$, 2) the N-formulation, in which $\bot$ was not postulated and $\neg$ was primitive, and the FN-formulation, in which both $\bot$ and $\neg$ were primitive and the equivalence of $\neg A$ and $A \supset \bot$ was a consequence of the logical rules.

Chapter 7 was on quantification, and while not much was new here, the chapter was quite complete. The tableau system of Chapter 5 was extended to all quantified systems.

Chapter 8 was on modal systems.

In his [1965], Curry gave an algorithm for converting the linear form of natural deduction rules as used in [Jaśkowski, 1934; Fitch, 1952] to the form based on trees used by Gentzen and Curry himself, discussed some improvements to his semantic justification of the L-rules, and introduced a form of dialogue or game logic based on the ideas of [Lorenz, 1961; Lorenzen, 1961; Stegmüller, 1964][63] equivalent to LJ^*.

In his [1968a], he showed how to associate formulas with sequents in such a way that for an L-system satisfying certain conditions, if one sequent can be derived from some others by the rules of the system, then a sequent can be proved in which the right-hand side is the formula corresponding to the concluding sequent and the formulas on the left are those corresponding to the premise sequents.

This was Curry's last paper on proof theory proper. However, since his work on proof theory was based in many ways on his idea of formal system, it is worth mentioning that in his [1968b], he proved that his notion of formal system was essentially equivalent to that of Smullyan [1961].

[63] The first two of these works were later reprinted in [Lorenzen and Lorenz, 1978].

4 CURRY'S PHILOSOPHY OF MATHEMATICS

As part of his earliest work on combinatory logic, Curry developed a philosophy of mathematics that he was to call *formalism*. Although he had already formulated this philosophy in a preliminary way by 1929[64] the first exposition of it came with his [1939b], and it is best known from his [1951], most of which was written in 1939. To this day, this is all many philosophers of mathematics know of Curry's philosophy, even though both of these works date from 1939, early in his career.[65]

For Curry, the basic idea of *formalism* was that mathematics does not have a subject matter that is independent of our knowledge and comes from outside the subject. This is different from the views of both platonists and intuitionists, who believe that mathematics does have a subject matter external to the subject: platonists believe that there are such things as numbers, points, lines, etc. which exist independently of our knowledge of them, and intuitionists believe that the subject matter of mathematics consists of mental constructions which are based on a special kind of primordial intuition.[66] For Curry, mathematics is like language, in the sense that it is the result of human activity, and in the terms used by Karl Popper [1968][67] it is part of the "third world." Thus, for Curry, no mathematical objects existed before human beings started doing mathematics.

In 1939, Curry proposed to define mathematics as the *science of formal systems*.[68] Many logicians think in terms of formal systems whose formal objects represent propositions. But other kinds of formal systems are possible. For example, consider the system $\mathcal{N}$:[69] there is one atom, 0, and one primitive operation, denoted by adding | on the right, so obs are represented by adding strings of the symbol | to the right of 0. Thus, the obs all have the form

$$0\underbrace{||\ldots|}_{n},$$

which represents the natural number n. There is one elementary predicate, which takes two arguments, so an elementary statement is written

$$X = Y$$

There is one axiom, namely

$$0 = 0,$$

[64] See my paper [1980, page 11].

[65] See, for example, [Shapiro, 2000, pages 168–170].

[66] See [Curry, 1963, page 9].

[67] This paper was delivered under the title "Epistemology and Scientific Knowledge" at a session of the Third International Congress for Logic, Methodology and Philosophy of Science in Amsterdam on Friday, August 25, 1967 at a session with Curry in the chair and at which I was present. After Popper finished speaking, Curry told me privately that he thought Popper had made a big deal out of something that was trivially true.

[68] See Definition 2 above.

[69] This is the system $\mathcal{N}_1$ of [Seldin, 1975a, page 456], which is the same as the system of [Curry, 1963, page 256].

and there is one rule:

$$X = Y \quad \Rightarrow \quad X| = Y|.$$

It is easy to see that all the obs represent natural numbers, and all the theorems have the form

$$X = X.$$

All of arithmetic and analysis can be constructed as part of the metatheory of this formal system $\mathcal{N}$, and by using the techniques of analytic geometry, it is also possible to make geometry part of this metatheory. This method of reconstructing mathematics in terms of formal systems seems different from the way most logicians and philosophers of mathematics think of using formal systems.

The definition of mathematics as the science of formal systems led to the criticism that he could not account for the existence of mathematics before formal systems first appeared at the end of the nineteenth century.[70] By the early 1960s, Curry proposed defining mathematics as the *science of formal methods*.[71] That this definition is broader than the earlier one can be seen by the fact that in elementary arithmetic, the operation of recognizing what a pile of five apples has in common with a pile of five oranges is a matter of abstracting from the kind of fruit one has in the pile and noting that what they have in common is their cardinality. Doing this is, in a sense, ignoring the *content* of the piles and noting only their *form*, which is a kind of formal activity.

To further clarify this idea, consider what Curry has to say in his [1963, page12] about natural numbers:

> A formalist would not speak of "the natural numbers" but of a set or system of natural numbers. Any system of objects, no matter what, which is generated from a certain initial object by a certain unary operation in such a way that each newly generated object is distinct from all those previously formed and that the process can be continued indefinitely, will do as a set of natural numbers. He may, and usually does, objectify this process by representing the numbers in terms of symbols; he chooses some symbol, let us say a vertical stroke '|', for the initial object, and regards the operation as the affixing of another '|' to the right of the given expression. But he realizes there are other interpretations; in particular, if one accepts the platonist or intuitionist metaphysics, their systems will do perfectly well.

Note that this "system of objects" need not be given by a formal system. A natural number is any object of a system of this kind.

In my opinion, this means that Curry's formalism, at least the formalism of his mature years, is a form of *structuralism*. To see this, note what Stewart Shapiro says in his [2000, page 258]:

[70] See, for example, [Shapiro, 2000, page 170], but this criticism was made much earlier.
[71] See [Curry, 1963, page 14].

> The structuralist vigorously rejects any sort of ontological independence among the natural numbers. The essence of a natural number is its *relations* to other natural numbers. The subject-matter of arithmetic is a single abstract structure, the pattern common to any infinite collection of objects that has a successor relation, a unique initial object, and satisfies the induction principle. The number 2 is no more and no less than the second position in the natural number structure; and 6 is the sixth position. Neither of them has any independence from the structure in which they are positions, and as positions in this structure, neither number is independent of the other. (Emphasis in original.)

What Curry would have said about this is that extracting the pattern to which Shapiro refers is a matter of abstracting the form from the structure and disregarding the content. And if a structuralist does not assume that these structures exists independently of our knowledge of them, Curry would have regarded that structuralist as his kind of formalist.

Appendix

A PURE λ-CALCULUS

In our discussion of functions at the beginning of this article, we mentioned two operations:

1. The *application* of a function to its argument, as in $f(3)$.
2. The *abstraction* operation that defines a function from an expression for its values, as in $\lambda x \,.\, x^2$.

These two operations are the basis of the λ-*calculus*, which is a formal system for functions interpreted as rules of calculation. The λ-calculus is the legacy of Alonzo Church.

The λ-calculus differs from ordinary mathematical practice in two essential ways:

1. Application is taken as a binary operation. At first glance, this may appear to leave out functions of more than one argument. But this is not so. There is a way to reduce functions of more than one argument to functions of one argument, a way called *currying* after Haskell B. Curry. Consider the function $f(x, y) = x - y$ of two arguments. If we replace x by some constant, say 4, we get $g(y) = f(4, y) = 4 - y$, a function of one argument. This function can be taken as the value of a function f^* related to f, so that $f^*(4) = g(y) = 4 - y$, and then $f^*(4)(y) = 4 - y$, and $f^*(4)(3) = 4 - 3 = 1$. This function f^* is called the *curried version of* f, or $\mathsf{curry}\, f$. Thus, $(\mathsf{curry}\, f)\, 4\, 3 = f(4, 3)$, and $(\mathsf{curry} f)\, x\, y = f(x, y)$.
2. In ordinary mathematical discourse, each function has a domain, and a function can only be applied to arguments from that domain. In pure λ-calculus, every term is considered to be in the domain of every other term, and every term is a function whose domain includes every other term. At first glance, this may seem counter-intuitive, and it does make models of λ-calculus very complicated. But it does correspond, in a sense, to what goes on in a computer, where the program instructions are stored in memory as is the data to which the program is applied, and all items stored in memory are simply strings of 0s and 1s.[72] In principle, there is no reason why a computer could not be told to take any such string as the program and any other such string as the data; if the first string is not in fact an executable set of instructions, the computer will not do anything useful, but only the human user will know that. The pure λ-calculus is simpler for not trying to define a domain for each function, and so the applicability of any term to any other term is not considered a problem.[73]

[72] Or, more precisely, pulses of high voltage and low voltage which are usually interpreted as 0s and 1s.

[73] Domains of λ-terms can be specified by using *types*. This is discussed in Appendix B.

This has to do with the formation of *terms*. But any formal system for functions as rules must deal with the computations involved in calculating the values of functions. In the discussion of the squaring function at the beginning of this article, we noted that to evaluate the function at an argument, we always start by substituting the argument for the variable in the expression for the function. Thus, we had that

$$(\lambda x \,.\, x^2)3 = 3^2.$$

In general, we would have

$$(\beta) \qquad (\lambda x \,.\, M)N = [N/x]M,$$

where $[N/x]M$ represents the result of substituting N for x in M. This is, in fact, the only means the λ-calculus has for carrying out calculations. It is perhaps surprising that it is enough to carry out all recursive operations.

Now the substitution mentioned in the last paragraph is not simple replacement, for we are dealing with *bound variables*. A bound variable is one for which no substitution is possible, and it serves only as a place-holder. Furthermore, along with computation rule (β), there is a rule which asserts that bound variables can always be changed:

(α) $(\lambda x \,.\, M) = (\lambda y \,.\, [y/x]M)$, if y does not occur free in M.

Finally, let us note an additional rule of computation that is often postulated. Note that if x does not occur free in M, and if N is any term, then

$$(\lambda x \,.\, Mx)N = MN.$$

This means that $(\lambda x \,.\, Mx)$ has the same effect on an argument N as M. So some systems of λ-calculus include the rule

(η) $(\lambda x \,.\, Mx) = M$, if x does not occur free in M.

Without this last rule, the system is called the $\lambda\beta$-calculus; with it, the calculus is called the $\lambda\beta\eta$-calculus or the $\lambda\eta$-calculus.

So far, the rules of calculation have been stated as equations. But in many actual calculation these rules are all read one way, from left to right. One calculation step that proceeds this way is called a *contraction*, and a sequence of such steps is called a *reduction*.

With these ideas, the pure λ-calculus is defined as follows:

Definition 3. *The formal system* Λ *of* λ-calculus *is defined as follows:*

1. *The* terms *are defined by:*

   (a) Variables. *There are assumed to be countably many, and it is assumed that they come in an enumerated sequence. Here, variables will be denoted by lower-case Italic letters from the end of the alphabet,* x, y, z, x_1, *etc.*

(b) Constants. *There may not be any, or there may be several, depending on the variant of system.*

(c) Applications. *If M and N are terms, then so is (MN). Parentheses are omitted by association to the left, so that $MN_1N_2 \dots N_n$ is $(\dots((MN_1)N_2)\dots N_n)$.*

(d) Abstractions. *If x is a variable and M is a term, then so is $(\lambda x \,.\, M)$. Repeated abstraction $(\lambda x_1 \,.\, (\lambda x_2 \,.\, \dots (\lambda x_n \,.\, M)\dots))$ will be written $(\lambda x_1 x_2 \dots x_n \,.\, M)$.*

2. *An occurrence of a variable x in a term M is said to be* bound *if it occurs in a subterm of M of the form $(\lambda x \,.\, N)$. An occurrence of a variable which is not bound is said to be* free. *That a variable x has a free occurrence in a term M is written $x \in \mathrm{FV}(M)$.*

3. *The* substitution *of a term N for a variable x in a term M, denoted by $[N/x]M$, is defined by induction on the structure of M as follows:*

   (a) *$[N/x]x$ is N.*

   (b) *If a is any variable different from x or is a constant, then $[N/x]a$ is a.*

   (c) *$[N/x](M_1M_2)$ is $([N/x]M_1[N/x]M_2)$.*

   (d) *$[N/x](\lambda x \,.\, M)$ is $(\lambda x \,.\, M)$.*

   (e) *If y is different from x, and if either $y \notin \mathrm{FV}(N)$ or else $x \notin \mathrm{FV}(M)$, then $[N/x](\lambda y \,.\, M)$ is $(\lambda y \,.\, [N/x]M)$.*

   (f) *If $y \in \mathrm{FV}(N)$ and $x \in \mathrm{FV}(M)$, then $[N/x](\lambda y \,.\, M)$ is defined to be $(\lambda z \,.\, [z/y][N/x]M)$, where z is the first variable in the given sequence of variables which is not in $\mathrm{FV}(MN)$.*

   (There are many other notations for $[N/x]M$, including $M[N/x]$ and $M[x := N]$.)

4. *There are two kinds of* reduction. *That M reduces to N will be written $M \rhd N$.*

   (a) *The first kind, known as β-*reduction, *is defined by the following axioms and rules:*

      (α) $(\lambda x \,.\, M) \rhd (\lambda y \,.\, [y/x]M)$, where $y \notin \mathrm{FV}(M)$.

      (β) $((\lambda x \,.\, M)N) \rhd [N/x]M$.

      (ρ) $M \rhd M$.

      (μ) If $M \rhd N$, then $PM \rhd PN$.

      (ν) If $M \rhd N$, then $MP \rhd NP$.

      (ξ) If $M \rhd N$, then $(\lambda x \,.\, M) \rhd (\lambda x \,.\, N)$.

      (τ) If $M \rhd N$ and $N \rhd P$, then $M \rhd P$.

This is equivalent to defining β-reductions as involving two kinds of replacements:

(α) If $y \notin \mathrm{FV}(M)$, a replacement of $(\lambda x \,.\, M)$ by $(\lambda y \,.\, [y/x]M)$.

(β) $((\lambda x \,.\, M)N)$ by $[N/x]M$. Here $((\lambda x \,.\, M)N)$ is called a β-redex, *and $[N/x]M$ is called its* contractum.

Each of these kinds of replacements is called a contraction.

The notation $M \rhd_\beta N$ will mean that M β-reduces to N.

(b) The second kind of reduction, $\beta\eta$-reduction, *is defined by adding to the definition of β-reduction the axiom*

(η) If $x \notin \mathrm{FV}(M)$, then $(\lambda x \,.\, Mx) \rhd M$.

This is equivalent to adding the replacement

(η) $(\lambda x \,.\, Mx)$ by M, provided that $x \notin \mathrm{FV}(M)$. Here $(\lambda x \,.\, Mx)$ is called an η-redex *and M is its* contractum.

This kind of replacement is called an η-contraction.

The notation $M \rhd_{\beta\eta} N$ means that M $\beta\eta$-reduces to N.

5. Conversion. *The relation $M =_* N$ (M* converts *to N), is defined by changing $\rhd$ to $=_*$ in the definition of reduction and adding the rule*

(σ) If $M =_* N$ then $N =_* M$.

Conversion can also be defined as a sequence of contractions and reverse contractions. There are two kinds of conversion: β-conversion, $=_\beta$, and $\beta\eta$-conversion, $=_{\beta\eta}$.

REMARK 5. *For $=_{\beta\eta}$, we have the following derived rule:*

(ζ) *If $Mx =_{\beta\eta} Nx$ and if $x \notin \mathrm{FV}(MN)$, then $M =_{\beta\eta} N$.*

This rule is also called (Ext), (for extensionality*).*

REMARK 6. *Terms which cannot be reduced (i.e., terms in which there are no occurrences of redexes), are said to be in* normal form.

REMARK 7. *In general, the symbol $\equiv$ will be used to indicate identical terms. In λ-calculus, however, it will mean terms that are (α)-equivalent (i.e., that differ only in their bound variables). This is because it is common in λ-calculus to identify terms that differ only in their bound variables, since they mean the same thing.*[74]

REMARK 8. *There is a fixed-point operator Y, which, when applied to any term, calculates a fixed-point of that term. In fact, there is more than one. One example is*

$$\mathsf{Y} \equiv (\lambda xz \,.\, x(zzx))(\lambda xz \,.\, x(zzx)).$$

[74]But sometimes ignoring the difference between distinct (α)-convertible terms can lead to technical difficulties.

It is easy to show that $\mathsf{Y}x \rhd x(\mathsf{Y}x)$.

REMARK 9. *Natural numbers and numerical functions can be coded in* Λ*, by a method originally due to Alonzo Church, as follows:*

- $\mathbf{0}$ *is defined to be* $\lambda xy\,.\,x$
- $\mathbf{1}$ *is defined to be* $\lambda xy\,.\,xy$.
- $\mathbf{2}$ *is defined to be* $\lambda xy\,.\,x(xy)$.
- $\vdots$
- $\mathbf{n}$ *is defined to be* $\lambda xy\,.\,\underbrace{x(\ldots(x(x}_{n}\,y))\ldots)$.

Numerical functions can be represented as terms of Λ*. A numerical function can be represented as a term in* Λ *if and only if it is recursive.*

REMARK 10. *Church did not define* $\mathbf{0}$ *in his system but started with* $\mathbf{1}$ *because he wanted* $\lambda x\,.\,M$ *to be well-formed only if there is a free occurrence of* x *in* M*, and so in his system* $\mathbf{0}$ *is not well-formed. This form of* λ*-calculus is called* λI*-calculus. See § 1.2 above. The system* Λ *defined here is often called* λK*-calculus when it needs to be distinguished from the* λI*-calculus.*

Since reduction represents the process of calculation, it is important to have confidence that the order in which calculation steps are performed does not make a difference. That it does not is a result of the *Church-Rosser Theorem*:

THEOREM 11 Church-Rosser Theorem. *If* $M \rhd P$ *and* $M \rhd Q$*, then there is a term* N *such that* $P \rhd N$ *and* $Q \rhd N$.

The theorem holds for both β-reduction and $\beta\eta$-reduction.

Corollary 1. If $M =_* N$, then there is a term P such that $M \rhd P$ and $N \rhd P$.

Here, if $=_*$ is $=_\beta$, then $\rhd$ is $\rhd_\beta$, whereas if $=_*$ is $=_{\beta\eta}$, then $\rhd$ is $\rhd_{\beta\eta}$.
For more details, see [Hindley and Seldin, 1986] and the references given there.

B LAMBDA-CALCULUS WITH TYPES

In pure λ-calculus, every term is in the domain of every other term. If it is necessary to distinguish domains of function, we use *types*. There are now many versions of type assignment to λ-terms, but the original and most basic one, which is the one we consider here, uses the simplest definition of types.

Definition 4. Types *are defined from* atomic types $\theta_1, \theta_2, \ldots, \theta_n, \ldots$ *(possibly infinitely many) by induction:*

1. *Every atomic type is a type.*
2. *If* α *and* β *are types, then so is* $(\alpha \to \beta)$.

REMARK 12. *Other notations have been used in the past for the type* $(\alpha \to \beta)$*, including* $(\alpha\beta)$*,* $(\beta\alpha)$*, and Curry's notation* $\mathsf{F}\alpha\beta$*. The notation* $(\alpha \to \beta)$ *is now standard. The outermost parentheses are often omitted. Also, parentheses are omitted by association to the right, so that*

$$\alpha_1 \to \alpha_2 \to \ldots \to \alpha_n \to \beta$$

is the type

$$(\alpha_1 \to (\alpha_2 \to (\ldots(\alpha_n \to \beta)\ldots))).$$

The type $(\alpha \to \beta)$ *is intended to represent the type of functions which take arguments of type* α *and have values in* β*.*

There are two approaches to the assignment of types to λ-terms, one due to Church and the other due to Curry.

In the Church style, every variable is assigned a unique type, and this type is indicated in λ-abstracts, so that if x has type α, the abstraction of M with respect to x is denoted $\lambda x : \alpha \,.\, M$. (An older notation, used by Church, is $\lambda x_\alpha \,.\, M_\beta$, where β is the type of M.) In Church-style typing, $\lambda x : \alpha \,.\, x$ (or, in Church's original notation $\lambda x_\alpha \,.\, x_\alpha$) is an identity function whose domain and range are the type α, and each type has its own identity function.

That a type α is assigned to a term M is denoted $M : \alpha$. If there are any atomic constants, types are assigned to them by axioms, but there may not be any atomic constants. The type assignment system, TACh_λ, is defined as follows:

Definition 5. *The* system TACh_λ *is defined as follows:*

1. Types *are defined as in Definition 4.*

2. Terms *are like the terms of* Λ*, except that clause 1d of Definition 3 is replaced by the following: if* x *is a variable,* α *is a type, and* M *is a term, then* $\lambda x : \alpha \,.\, M$ *is a term.*

3. *Typing judgements are all of the form* $M : \alpha$*, where* M *is a term and* α *is a type.*

4. *If* c *is an atomic (term) constant with a type* γ*, then*

$$c : \gamma$$

is an axiom. (That this is an axiom is often written

$$\overline{c : \gamma},$$

to indicate a step of a deduction with no premises.)

5. *Types are assigned to compound terms by the following two rules:*

$(\to E)$ $$\frac{M : \alpha \to \beta \quad N : \alpha}{MN : \beta}$$

$(\to I_{\mathrm{Ch}})$ $$\frac{\begin{array}{c}[x : \alpha] \\ M : \beta\end{array}}{\lambda x : \alpha \,.\, M : \alpha \to \beta}$$

Condition: *x does not occur free in any undischarged assumption.*

In Curry-style typing, the types of the variables are not fixed, but can be any types for which the rules allow a term to be assigned a type. The terms are simply the terms of Λ. In Curry-style typing, $\lambda x \,.\, x$ is an identity function on any type, so $\lambda x \,.\, x : \alpha \to \alpha$ for every type α.

The formal definition is as follows

Definition 6. *The* system TACu_λ *is defined as in Definition 5 with the following changes:*

1. *In Clause 2 of that definition, the* terms *are the terms of* Λ.
2. *In Clause 5 of that definition, rule* $(\to I_{\mathrm{ch}})$ *is replaced by*

$(\to I_{\mathrm{Cu}})$ $$\frac{\begin{array}{c}[x : \alpha] \\ M : \beta\end{array}}{\lambda x \,.\, M : \alpha \to \beta}$$

Condition: *x does not occur free in any undischarged assumption.*

Note that in both styles of typing, the deduction of a type assignment follows the construction of the term.

These rules are for deductions from assumptions. Assumptions are usually assumed in the form of *contexts*, which assign variables to types, assigning distinct variables to distinct types. Thus, a context would have the form

$$x_1 : \alpha_1, x_2 : \alpha_2, \ldots, x_n : \alpha_n.$$

Here, the variables $x_1, x_2, \ldots, x_n$ are all required to be distinct (but the types $\alpha_1, \alpha_2, \ldots, \alpha_n$ need not be distinct).

Some important results about this system are the following:

THEOREM 13 Subject-Reduction Theorem. *If* $M \rhd N$ *and* $\Gamma \vdash M : \alpha$ *for any context* Γ *and any type* α*, then* $\Gamma \vdash N : \alpha$.

THEOREM 14 Normal Form Theorem. *If* $\Gamma \vdash M : \alpha$ *for any context* Γ *and any type* α*, then* M *has a normal form.*

C PURE COMBINATORY LOGIC

The λ-calculus has two primitive operations for constructing terms: abstraction and application. In *combinatory logic*, there is only one such operation, application. Instead of abstraction, combinatory logic has some special constants, called *basic combinators*, and from these basic combinators abstraction can be defined. Since the only primitive term-forming operation is application, there are no bound variables to worry about, and variables will simply be atomic terms that we decide to call variables and for which we are prepared to make substitutions.[75] Furthermore, the substitution of a term Y for a variable x in another term X is simply the replacement of all occurrences of x by Y.

In standard combinatory logic, the basic combinators are K and S. These two constants have reduction rules defined as follows:

(K) $\mathsf{K}XY \rhd X$,

(S) $\mathsf{S}XYZ \rhd XZ(YZ)$.

(Weak) reduction, or $\rhd_w$ is defined as replacements using these two rules, and the two kinds of redexes are of the form $\mathsf{K}XY$ and $\mathsf{S}XYZ$. Its corresponding conversion is $=_w$. It is easy to see that if $X \rhd Y$, then for any term Z, $[Z/x]X \rhd [Z/x]Y$.

Other important functions can be defined in terms of these two, for example the identity function I is defined to be SKK, since

$$\mathsf{SKK}X \rhd \mathsf{K}X(\mathsf{K}X) \rhd X.$$

Other definitions are for the composition operator B, which is defined to be $\mathsf{S}(\mathsf{KS})\mathsf{K}$ and has the property that $\mathsf{B}XYZ \rhd X(YZ)$, the commutator C, which has the property that $\mathsf{C}XYZ \rhd XZY$ and is defined to be $\mathsf{S}(\mathsf{BBS})(\mathsf{KK})$, and the duplicator W, which satisfies $\mathsf{W}XY \rhd XYY$ and is defined to be $\mathsf{SS}(\mathsf{KI})$.

It turns out that the combinator S can be defined in terms of B, C, and W as $\mathsf{B}(\mathsf{B}(\mathsf{BW})\mathsf{C})(\mathsf{BB})$; see [Curry and Feys, 1958, §5B1].

In his dissertation [1935a; 1935b], Rosser discovered that all the other combinators can be defined in terms of I and a combinator J with the reduction rule

$$\mathsf{J}UXYZ \rhd UX(UZY).$$

Now, for a variable x, an abstraction term $[x]X$ can be defined by induction on the structure of X as follows:

- If a is an atomic term, then $[x]a$ is
  - I if a is x,

[75] This is essentially the same procedure used in computer languages, where variables are simply identifiers for which we are prepared to make substitutions and constants are simply identifiers for which we are not prepared to make substitutions.

– $\mathsf{K}a$ if a is not x.

- If X is YZ, then assuming that $[x]Y$ and $[x]Z$ are already defined, $[x]XY$ is $\mathsf{S}([x]Y)([x]Z)$.

Then it follows that $[x]X$ is a term in which the variable x does not occur. Furthermore, it can easily be proved by induction on the structure of the term X that

$$([x]X)Y \rhd [Y/x]X.$$

However, this is not the only way to define abstraction, and this definition results in longer terms than some others. For example, the following definition is often preferred:

(a) If $x \notin \mathrm{FV}(U)$, then $[x]U$ is $\mathsf{K}U$,

(b) $[x]x$ is I, and

(f) If X is YZ, $x \in \mathrm{FV}(X)$, and $[x]Y$ and $[x]Z$ are already defined, then $[x]X$ is $\mathsf{S}([x]Y)([x]Z)$.[76]

Still other definitions are possible, and some are significantly more efficient than this one.

REMARK 15. *The notation $[x]X$ is Curry's original notation, and it is still in use by many computer scientists. The abstraction operator in combinatory logic is often called "bracket abstraction." Another notation, proposed by Barendregt [Barendregt, 1984] and used in [Hindley and Seldin, 1986] is $\lambda^* x \,.\, X$, but this notation seems currently less common.*

We can now formally define combinatory logic:

Definition 7. *The formal system H of* combinatory logic *is defined as follows:*

1. Terms *are defined as follows:*
   - *(a)* Variables. *There is given an infinite sequence of variables, x, y, z, x_1, etc.*
   - *(b)* Constants. *There are two special constants:* K *and* S. *There may be more constants.*
   - *(c)* Applications. *If X and Y are terms, so is (XY).*
2. *All occurrences of variables in terms are free. That x occurs in X will be denoted $x \in \mathrm{FV}(X)$.*
3. *The* substitution *of a term Y for a variable x in another term X, $[Y/x]X$, is defined as in Definition 3, Clause 3, subclauses (a)-(c).*

[76] On the names of these clauses, see [Curry and Feys, 1958, §6A3, page 190] and [Curry *et al.*, 1972, §11C1, page 43].

4. Abstraction *is defined by the second definition given above just before Remark 15.*

5. (Weak) reduction, $\rhd_w$, *is defined by the following axioms and rules:*

   (K*)* $\mathsf{K}XY \rhd X$.

   (S*)* $\mathsf{S}XYZ \rhd XZ(YZ)$.

   (ρ*)* $X \rhd X$.

   (μ*) If* $X \rhd Y$, *then* $ZX \rhd ZY$.

   (ν*) If* $X \rhd Y$, *then* $XZ \rhd YZ$.

   (τ*) If* $X \rhd Y$ *and* $Y \rhd Z$, *then* $X \rhd Z$.

   This is equivalent to defining weak reductions as replacements of redexes by contracta, where there are two kinds of redexes: 1) K *redexes (*$\mathsf{K}XY$*), whose contractum is* X, *and 2)* S *redexes (*$\mathsf{S}XYZ$*), whose contractum is* $XZ(YZ)$.

6. (Weak) conversion, $=_w$, *is defined by changing* $\rhd$ *to* $=_w$ *in the definition of weak reduction and adding the rule*

   (σ*) If* $X =_w Y$, *then* $Y =_w X$.

   Weak reduction can also be defined as a sequence of contractions and reverse contractions.

REMARK 16. *Weak reduction and conversion satisfy the Church-Rosser Theorem, i.e., Theorem 11 and Corollary 1.*

Combinatory terms can be translated to λ-terms by the following translation:

Definition 8. *The* translation $-_L$ *from terms in H to terms in* Λ *is defined as follows:*

1. *If* a *is any atomic term except* K *or* S, *then* $a_L \equiv a$.

2. $\mathsf{K}_L \equiv \lambda xy\,.\,x$.

3. $\mathsf{S}_L \equiv \lambda xyz\,.\,xz(yz)$.

4. $(XY)_L \equiv (X_L Y_L)$.

With this definition, it is easy to prove the following:

THEOREM 17. *If* $X \rhd_w Y$ *in H, then* $X_L \rhd_\beta Y_L$ *in* Λ. *If* $X =_w Y$ *in H, then* $X_L =_\beta Y_L$.

The converse translation depends on the abstraction algorithm used in H:

Definition 9. *The* translation $-_H$ *from* Λ *to H is defined as follows:*

1. *If a is an atomic term, then $a_H \equiv a$.*
2. *$(MN)_H \equiv (M_H N_H)$.*
3. *$(\lambda x\ .\ M)_H \equiv ([x]M_H)$.*

REMARK 18. *Note that for any abstraction algorithm, if $y \notin \mathrm{FV}(M)$, then $([x]M)_H \equiv ([y][y/x]M_H)$. Thus, rule (α) is mapped to an identity between terms by the mapping $-_H$. This is another reason for identifying α-convertible terms in λ-calculus.*

REMARK 19. *There is no theorem for the translation $-_H$ corresponding to Theorem 17. The reason for this is that rule (ξ) does not hold for $\rhd_w$. This can be seen from the following example: $\mathsf{I}x \rhd_w x$, but*

$$[x](\mathsf{I}x) \equiv \mathsf{S}(\mathsf{KI})\mathsf{I} \not\rhd_w \mathsf{I} \equiv [x]x.$$

A form of $\rhd_w$ and $=_w$ can be defined for λ-terms: drop rule (ξ) or, which is equivalent, forbid replacements within the scope of a λ-abstract. For this reduction, we do have that if $M \rhd_w N$ (respectively $M =_w N$), then $M_H \rhd_w N_H$ (respectively $M_H =_w N_H$).

D COMBINATORS WITH TYPES

To have a Church-style typing, we would need a combinator $\mathsf{K}_{\alpha\beta}$ of type $\alpha \to \beta \to \alpha$ for each pair of types α and β, and similarly for S, so that there would be a combinator I_α of type $\alpha \to \alpha$ for each type α. This would lead to an infinite number of atomic combinators, and so is not usually considered. Instead, type assignment for combinators is usually taken in the Curry style.

In assigning types to terms of combinatory logic, deductions that terms have types will follow the construction of the terms. For this reason, the definition of the system is as follows:

Definition 10. *The* system TACu *is defined as follows:*

1. *The* types *are those defined in Definition 4 above.*
2. *The* terms *are the terms of the system H of Definition 7 above.*
3. *Typing judgements are of the form $X : \alpha$, where X is a term and α is a type.*
4. *If c is an atomic constant different from K and S, and if c is assigned type γ, then $c : \gamma$ is an axiom. There are two additional axiom schemes:*

   (FK) For all types α and β, $\mathsf{K} : (\alpha \to \beta \to \alpha)$.

   (FS) For all types α, β, and γ, $\mathsf{S} : (\alpha \to \beta \to \gamma) \to (\alpha \to \beta) \to (\alpha \to \gamma)$.
5. *The rule of the system is the rule ($\to$ E) of Definition 5, clause 5.*

The Subject-Reduction Theorem, Theorem 13 above, and the Normal Form Theorem, Theorem 14 above, both hold for system TACu.

BIBLIOGRAPHY

[Andrews, 1965] P. B. Andrews. *A Transfinite Type Theory with Type Variables*. North-Holland Publishing Company, Amsterdam, 1965.

[Andrews, 1986] P. B. Andrews. *An Introduction to Mathematical Logic and Type Theory: to Truth thorugh Proof*. Academic Press, 1986.

[Barendregt *et al.*, 1993] H. P. Barendregt, M. W. Bunder, and W. J. W. Dekkers. Systems of illative combiantory logic complete for first-order propositional and predicate calculus. *Journal of Symbolic Logic*, 58(3):769–788, September 1993.

[Barendregt, 1984] H. P. Barendregt. *The Lambda Calculus: Its Syntax and Semantics*. North-Holland Publishing Company, Amsterdam, New York, and Oxford, 1984.

[Bernays, 1953] P. Bernays. Review of H. B. Curry, "The system LD," J. Symbolic Logic 17:35–42, 1952 and "On the definition of negation by a fixed proposition in the inferential calculus," Ibid., pp. 98–104. *Journal of Symbolic Logic*, 18:266–268, 1953.

[Bunder and Meyer, 1978] M. W. Bunder and R. K. Meyer. On the inconsistency of systems similar to $\mathcal{F}^*_{21}$. *Journal of Symbolic Logic*, 43(1):1–2, March 1978.

[Bunder, 1969] M. W. Bunder. *Set Theory Based on Combinatory Logic*. PhD thesis, University of Amsterdam, 1969.

[Bunder, 1974] M. W. Bunder. Some inconsistencies in illative combinatory logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 20:199–201, 1974.

[Bunder, 1976] M. W. Bunder. The inconsistency of $\mathcal{F}^*_{21}$. *Journal of Symbolic Logic*, 41(2):467–468, June 1976.

[Cardone and Hindley, 2008] F. Cardone and J. R. Hindley. History of lambda-calculus and combinatory logic. In this Volume, pp. 535–630.

[Church and Kleene, 1936] A. Church and S. C. Kleene. Formal definitions in the theory of ordinal numbers. *Fundamenta Mathematicae*, 28:11–31, 1936.

[Church and Rosser, 1936] A. Church and J. B. Rosser. Some properties of conversion. *Transactions of the American Mathematical Society*, 39:472–482, 1936.

[Church, 1924] A. Church. Uniqueness of the Lorentz transformation. *American Mathematical Monthly*, 31:376–382, 1924.

[Church, 1925] A. Church. On irredundant sets of postulates. *Transactions of the American Mathematical Society*, 27:318–328, 1925.

[Church, 1927] A. Church. Alternatives to Zermelo's assumption. *Transactions of the American Mathematical Society*, 29:178–208, January 1927.

[Church, 1928] A. Church. On the law of the excluded middle. *Bulletin of the American Mathematical Society*, 34:75–78, January–February 1928.

[Church, 1932] A. Church. A set of postulates for the foundation of logic. *Annals of Mathematics*, 33:346–366, 1932.

[Church, 1933] A. Church. A set of postulates for the foundation of logic (second paper). *Annals of Mathematics*, 34:839–864, 1933.

[Church, 1935] A. Church. A proof of freedom from contradiction. *Proceedings of the National Academy of Sciences of the United States of America*, 21:275–281, 1935.

[Church, 1936a] A. Church. Correction to *A note on the Entscheidungsproblem*. *Journal of Symbolic Logic*, 1(3):101–102, September 1936.

[Church, 1936b] A. Church. A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1(1):40–41, March 1936.

[Church, 1936c] A. Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58:345–363, 1936.

[Church, 1938] A. Church. The constructive second number class. *Bulletin of the American Mathematical Society*, 44:224–232, April 1938.

[Church, 1940] A. Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5:56–68, 1940.

[Church, 1941] A. Church. *The Calculi of Lambda Conversion*. Princeton University Press, 1941.

[Church, 1951] A. Church. A formulation of the logic of sense and denotation. In P. Henley, editor, *Sturcture, Method and Meaning: Essays in Honor of Henry M. Sheffer*, pages 3–24. Liberal Arts Press, New York, 1951.

[Church, 1973] A. Church. Outline of a revised formulation of the logic of sense and denotation (Part I). *Noûs*, 7(1):24–33, March 1973.

[Church, 1974] A. Church. Outline of a revised formulation of the logic of sense and denotation (Part II). *Noûs*, 8(2):135–156, May 1974.

[Church, 1989] A. Church. Intensionality and the paradox of the name relation. In J. Almog, J. Perry, and H. Wettstein, editors, *Themes from Kaplan*, pages 151–165. Oxford University Press, New York, 1989.

[Church, 1993] A. Church. A revised formulation of the logic of sense and denotation. Alternative (1). *Noûs*, 27(2):141–157, June 1993.

[Cogan, 1955] E. J. Cogan. A formalization of the theory of sets from the point of view of combinatory logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 1:198–240, 1955. PhD thesis, The Pennsylvania State University, 1955.

[Constable and others, 1986] R. Constable et al. *Implementing Mathematics with the Nuprl Proof Development System*. Prentice-Hall, Englewood Cliffs, NJ, 1986.

[Coquand and Huet, 1988] Thierry Coquand and Gérard Huet. The calculus of constructions. *Information and Computation*, 76:95–120, 1988.

[Curry and Feys, 1958] H. B. Curry and R. Feys. *Combinatory Logic*, volume 1. North-Holland Publishing Company, Amsterdam, 1958. Reprinted 1968 and 1974.

[Curry and Schoenberg, 1947] H. B. Curry and I. J. Schoenberg. On spline distributions and their limits: The Pólya distribution functions (Absract). *Bulletin of the American Mathematical Society*, 53:1114, 1947.

[Curry and Schoenberg, 1966] H. B. Curry and I. J. Schoenberg. On Pólya frequency functions IV: the fundamental spline functions and their limits. *Journal D'Analyse Mathématique*, 17:71–107, 1966.

[Curry *et al.*, 1972] H. B. Curry, J. R. Hindley, and J. P. Seldin. *Combinatory Logic*, volume 2. North-Holland Publishing Company, Amsterdam and London, 1972.

[Curry, 1929] H. B. Curry. An analysis of logical substitution. *American Journal of Mathematics*, 51:363–384, 1929.

[Curry, 1930] H. B. Curry. Grundlagen der kombinatorischen Logik. *American Journal of Mathematics*, 52:509–536, 789–834, 1930. Inauguraldissertation.

[Curry, 1931] H. B. Curry. The universal quantifier in combinatory logic. *Annals of Mathematics*, 32:154–180, 1931.

[Curry, 1932] H. B. Curry. Some additions to the theory of combinators. *American Journal of Mathematics*, 54:551–558, 1932.

[Curry, 1933] H. B. Curry. Apparent variables from the standpoint of combinatory logic. *Annals of Mathematics*, 34:381–404, 1933.

[Curry, 1934a] H. B. Curry. Functionality in combinatory logic. *Proceedings of the National Academy of Sciences of the United States of America*, 20:584–590, 1934.

[Curry, 1934b] H. B. Curry. Some properties of equality and implication in combinatory logic. *Annals of Mathematics*, 35:849–850, 1934.

[Curry, 1935] H. B. Curry. Foundations of the theory of abstract sets from the standpoint of combinatory logic (abstract). *Bulletin of the American Mathematical Society*, 40:654, 1935.

[Curry, 1936] H. B. Curry. First properties of functionality in combinatory logic. *Tôhoku Mathematical Journal*, 41:371–401, 1936.

[Curry, 1937] H. B. Curry. Some properties of formal deducibility (Abstract). *Bulletin of the American Mathematical Society*, 43:615, 1937.

[Curry, 1939a] H. B. Curry. A note on the reduction of Gentzen's calculus LJ. *Bulletin of the American Mathematical Society*, 45:288–293, 1939.

[Curry, 1939b] H. B. Curry. Remarks on the definition and nature of mathematics. *Journal of Unified Science*, 9:164–169, 1939. All copies of this issue were destroyed during World War II, but some copies were distributed at the International Congress for the Unity of Science in Cambridge, MA, in September, 1939. Reprinted with minor corrections in *Dialectica* 8:228–333, 1954. Reprinted again in Paul Benacerraf and Hilary Putnam (eds) *Philosophy of Mathematics: Selected Readings*, pages 152–156, Prentice Hall, Englewood-Cliffs, New Jersey, 1964.

[Curry, 1941a] H. B. Curry. The consistency and completeness of the theory of combinators. *Journal of Symbolic Logic*, 6:54–61, 1941.

[Curry, 1941b] H. B. Curry. The paradox of Kleene and Rosser. *Transactions of the American Mathematical Society*, 50:454–516, 1941.

[Curry, 1941c] H. B. Curry. A revision of the fundamental rules of combinatory logic. *Journal of Symbolic Logic*, 6:41–53, 1941.
[Curry, 1941d] H. B. Curry. Some aspects of the problem of mathematical rigor. *Bulletin of the American Mathematical Society*, 47:221–241, 1941.
[Curry, 1942a] H. B. Curry. The combinatory foundations of mathematical logic. *Journal of Symbolic Logic*, 7:49–64, 1942.
[Curry, 1942b] H. B. Curry. The inconsistency of certain formal logics. *Journal of Symbolic Logic*, 7:115–117, 1942.
[Curry, 1942c] H. B. Curry. Some advances in the combinatory theory of quantification. *Proceedings of the National Academy of Sciences of the United States of America*, 28:564–569, 1942.
[Curry, 1949] H. B. Curry. A simplification of the theory of combinators. *Synthese*, 7:391–399, 1949.
[Curry, 1950] H. B. Curry. *A Theory of Formal Deducibility*. Notre Dame Mathematical Lectures Number 6. University of Notre Dame, Notre Dame, Indiana, 1950. Second edition, 1957.
[Curry, 1951] H. B. Curry. *Outlines of a Formalist Philosophy of Mathematics*. North-Holland Publishing Company, Amsterdam, 1951.
[Curry, 1952a] H. B. Curry. The elimination theorem when modality is present. *Journal of Symbolic Logic*, 17(4):249–265, December 1952.
[Curry, 1952b] H. B. Curry. *Leçons de Logique Algébrique*. Gauthier-Villars and Nauwelaerts, Paris and Louvain, 1952.
[Curry, 1952c] H. B. Curry. A new proof of the Church-Rosser Theorem. *Indagationes Mathematicae*, 14:16–23, 1952.
[Curry, 1952d] H. B. Curry. On the definition of negation by a fixed proposition in inferential calculus. *Journal of Symbolic Logic*, 17(2):98–104, June 1952.
[Curry, 1952e] H. B. Curry. The permutability of rules in the classical inferential calculus. *Journal of Symbolic Logic*, 17(4):245–248, December 1952.
[Curry, 1952f] H. B. Curry. The system LD. *Journal of Symbolic Logic*, 17(1):35–42, March 1952.
[Curry, 1954] H. B. Curry. The logic of program composition. In *Applications Scientifiques de la Lagique Mathématique, Actes du* 2^e *Colloque International de Logique Mathématique, Paris—25–30 Août 1952, Institut Henri Poincaré*, pages 97–102. Gauthier-Villars, Paris, 1954.
[Curry, 1955] H. B. Curry. The inconsistency of the full theory of combinatory functionality (abstract). *Journal of Symbolic Logic*, 20:91, 1955.
[Curry, 1956] H. B. Curry. Consistency of the theory of functionality (abstract). *Journal of Symbolic Logic*, 21:110, 1956.
[Curry, 1960] H. B. Curry. The deduction theorem in the combinatory theory of restricted generality. *Logique et Analyse*, 3:15–39, 1960.
[Curry, 1961] H. B. Curry. Basic verifiability in the combinatory theory of restricted generality. In *Essays on the Foundations of Mathematics*, pages 165–189. Magnes Press, Hebrew University, Jerusalem, 1961.
[Curry, 1963] H. B. Curry. *Foundations of Mathematical Logic*. McGraw-Hill Book Company, Inc., New York, San Francisco, Toronto, and London, 1963. Reprinted by Dover, 1977 and 1984.
[Curry, 1965] H. B. Curry. Remarks on inferential deduction. In A.-T. Tymieniecka, editor, *Contributions to logic and Methodology in Honor of J. M. Bocheński*, pages 45–72. North-Holland, Amsterdam, 1965.
[Curry, 1968a] H. B. Curry. A deduction theorem for infrential predicate calculus. In K. Schütte, editor, *Contributions to Mathematical Logic*, pages 91–108. North-Holland, Amsterdam, 1968.
[Curry, 1968b] H. B. Curry. The equivalence of two definitions of elementary formal system. *Compositio Mathematica*, 20:13–20, 1968. Festschrift in honor of A. Heyting, edited by D. van Dalen, S. C. Kleene, and A. S. Troelstra, Wolters-Noordhoff, Groningen.
[Curry, 1969a] H. B. Curry. Modified basic functionality in combinatory logic. *Dialectica*, 23:83–92, 1969.
[Curry, 1969b] H. B. Curry. The undecidability of $\lambda\mathsf{K}$-conversion. In J. J. Bulloff, T. C. Holyoke, and S. W. Hahn, editors, *Foundations of Mathematics: Symposium Papers Commemorating the Sixtieth birthday of Kurt Gödel*, pages 10–14. Springer Verlag, Berlin, 1969.

[Curry, 1973] H. B. Curry. The consistency of a system of combinatory restricted generality. *Journal of Symbolic Logic*, 38:83–92, 1973.

[Curry, 1975a] H. B. Curry. Representation of Markov algorithms by combinators. In A. R. Anderson, R. B. Marcus, and R. A. Martin, editors, *The Logical Enterprise*, pages 109–119. Yale University Press, New Haven, Connecticut, 1975. Festschrift in honor of F. B. Fitch.

[Curry, 1975b] H. B. Curry. A study of generalized standardization in combinatory logic. In J. Diller and G. H. Müller, editors, *ISLIC Proof Theory Symposium. Dedicated to Kurt Schütte on the Occasion of his 65th Birthday. Proceedings of the International Summer Institute and Logic Colloquium, Kiel, 1974*, pages 44–55. Springer Verlag, Berlin, 1975. *Lecture Notes in Mathematics*, volume 500.

[Curry, 1979] H. B. Curry. On a polynomial representation of $\lambda\beta$-normal forms. In K. Lorenz, editor, *Konstruktionen versus Positionen: Beiträge zur Diskussion um die Konstruktive Wissenschaftstheorie*, pages 94–98. Walter de Gruyter, Berlin and New York, 1979.

[Curry, 1980] H. B. Curry. Some philosophical aspects of combinatory logic. In John Barwise, H. Jerome Keisler, and Kenneth Kunen, editors, *The Kleene Symposium: A Collection of Papers on Recursion Theory and Intuitionism*, pages 85–101. North-Holland, Amsterdam, 1980.

[Dekkers *et al.*, 1998a] W. J. M. Dekkers, M. W. Bunder, and H. P. Barendregt. Completeness of the propositions-as-types interpretation of intuitionistic logic into illative combinatory logic. *Journal of Symbolic Logic*, 63(3):869–890, September 1998.

[Dekkers *et al.*, 1998b] W. J. M. Dekkers, M. W. Bunder, and H. P. Barendregt. Completeness of two systems of illative combinatory logic for first-order propositional and predicate calculus. *Archive for Mathematical Logic*, 37(5–6):327–341, 1998.

[Enderton, 1995] H. B. Enderton. In memoriam: Alonzo Church. *Bulletin of Symbolic Logic*, 1(4):486–488, December 1995.

[Fitch, 1952] Fredric Brenton Fitch. *Symbolic Logic*. The Ronald Press Company, New York, 1952.

[Gandy, 1980] R. O. Gandy. An Early Proof of Normalization by A.M. Turing. In J. P. Seldin and J. R. Hindley. editors, *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, pages 453–455. Academic Press, 1980.

[Gentzen, 1934] Gerhard Gentzen. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1934. Translated in Szabo (ed.), *The Collected Papers of Gerhard Gentzen* as "Investigations into Logical Deduction," pages 68–131, North-Holland, Amsterdam and London, 1969.

[Girard, 1971] J.-Y. Girard. Une extension de l'interprétation de Gödel à l'analyse, et son application à l'élimination des coupures dans l'analyse et la théorie des types. In J. E. Fenstad, editor, *Proceedings of the Second Scandinavian Logic Symposium*, pages 63–92, Amsterdam, 1971. North-Holland.

[Girard, 1972] J.-Y. Girard. *Interprétation Fonctionnelle et Élimination des Coupures de L'Arithmétique d'Orde Supérieur*. PhD thesis, University of Paris VII, 1972.

[Gödel, 1940] Kurt Gödel. *The Consistency of the Continuum Hypothesis*. Princeton University Press, Princeton, New Jersey, USA, 1940.

[Goldstine, 1946] A. Goldstine. Report on the ENIAC (Electronic Numeric Integrator and Computer). Technical report, Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, 1946.

[Henkin, 1950] L. Henkin. Completeness in the theory of types. *Journal of Symbolic Logic*, 15(2):81–91, June 1950.

[Henkin, 1963] L. Henkin. A theory of propositional types. *Fundamenta Mathematicae*, 52:323–344, 1963.

[Hindley and Seldin, 1986] J. R. Hindley and J. P. Seldin. *Introduction to Combinators and λ-Calculus*. Cambridge University Press, 1986.

[Hindley, 1964] J. R. Hindley. *The Church-Rosser Property and a Result in Combinatory Logic*. PhD thesis, Newcastle upon Tyne, 1964.

[Hindley, 1969] J. R. Hindley. The principal type scheme of an object in combinatory logic. *Transactions of the American Mathematical Society*, 146:29–60, 1969.

[Howard, 1980] W. A. Howard. The formulae-as-types notion of construction. In J. Roger Hindley and Jonathan P. Seldin, editors, *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, pages 479–490. Academic Press, New York, 1980. A version of this paper was privately circulated in 1969.

[Jaśkowski, 1934] Stanisław Jaśkowski. On the rules of supposition in formal logic. *Studia Logica*, 1:5–32, 1934.
[Kamareddine *et al.*, 2004] F. Kamareddine, T. Laan, and R. Nederpelt. *A Modern Perspective on Type Theory*. Kluwer, Dordrecht, Boston, London, 2004.
[Kanger, 1955] S. Kanger. A note on partial postulate sets for propositional logic. *Theoria (Sweden)*, 21:99–104, 1955.
[Kleene and Rosser, 1935] S. C. Kleene and J. B. Rosser. The inconsistency of certain formal logics. *Annals of Mathematics*, 36:630–636, 1935.
[Kleene, 1934] S. C. Kleene. Proof by cases in formal logic. *Annals of Mathematics*, 35(3):529–544, July 1934.
[Kleene, 1935] S. C. Kleene. A theory of positive integers in formal logic. *American Journal of Mathematics*, 57:153–173, 219–244, 1935.
[Kleene, 1936a] S. C. Kleene. General recursive functions of natural numbers. *Mathematische Annalen*, 112:727–742, 1936.
[Kleene, 1936b] S. C. Kleene. λ-definability and recursiveness. *Duke Mathematical Journal*, 2:340–353, 1936.
[Kleene, 1938] S. C. Kleene. On notation for ordinal numbers. *Journal of Symbolic Logic*, 3(4):150–155, December 1938.
[Kleene, 1944] S. C. Kleene. On the forms of predicates in the theory of constructive ordinals. *American Journal of Mathematics*, 66:41–58, 1944.
[Kleene, 1955] S. C. Kleene. On the forms of predicates in the theory of constructive ordinals (second paper). *American Journal of Mathematics*, 77:405–428, 1955.
[Kleene, 1962] S. C. Kleene. Lambda-definable functionals of finite type. *Fundamenta Mathematicae*, 50:281–303, 1962.
[Kleene, 1981] S. C. Kleene. Origins of recursive function theory. *Annals of the History of Computing*, 3:52–67, 1981.
[Kripke, 1958] S. A. Kripke. The system LE. Submitted to Westinghouse Science Talent Search, February, 1958; not published, 1958.
[Lercher, 1963] B. Lercher. *Strong reduction and recursion in combinatory logic*. PhD thesis, The Pennsylvania State University, 1963.
[Loewen, 1962] K. Loewen. *A study of strong reduction in combinatory logic*. PhD thesis, the Pennsylvania State University, 1962.
[Lorenz, 1961] K. Lorenz. *Arithmetik und Logik als Spiele*. PhD thesis, Universität Kiel, 1961.
[Lorenzen and Lorenz, 1978] P. Lorenzen and K. Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1978.
[Lorenzen, 1961] P. Lorenzen. Ein dialogisches Konstruktivitätskriterium. In *Infinitistic Methods: Proceedings of the Symposium on Foundations of Mathematics, Warsaw, 2–9 September 1959*, pages 193–200. Pergamon Press and Państwowe Wydawnictwo Naukowe, 1961.
[Martin-Löf, 1975] P. Martin-Löf. An intuitionistic theory of types: Predicative part. In H. E. Rose and J. C. Shepherdson, editors, *Logic Colloquium '73*, pages 73–118. North-Holland Publishing Company, Amsterdam, 1975.
[Martin-Löf, 1984] P. Martin-Löf. Intuitionistic type theory. Bibliopolis, Naples, 1984. Notes by Giovanni Sambin of a series of lectures given in Padua, June 1980.
[Mezghiche, 1984] Mohamed Mezghiche. Une nouvelle Cβ-réduction dans la logique combinatoire. *Theoretical Computer Science*, 31:151–165, 1984.
[Mezghiche, 1989] Mohamed Mezghiche. On pseudo-cβ normal form in combinatory logic. *Theoretical Computer Science*, 66:323–331, 1989.
[Mezghiche, 1997] Mohamed Mezghiche. $c\beta$-machine with $\lambda\beta$-reduction. *Theoretical Computer Science*, 189:221–228, 1997.
[Newman, 1942] M. H. A. Newman. On theories with a combinatorial definition of "equivalence". *Annals of Mathematics*, 43(2):223–243, 1942.
[Popper, 1968] K. R. Popper. Epistemology without a knowing subject. In *Logic, Methodology and Philosophy of Science III: Proceedings of the Third International Congress for Logic, Methodology and Philosophy of Science, Amsterdam 1967*, pages 333–373, Amsterdam, 1968. North-Holland.
[Prawitz, 1965] D. Prawitz. *Natural Deduction*. Almqvist & Wiksell, Stockholm, Göteborg, and Uppsala, 1965. Reprinted by Dover, 2006.
[Rosser, 1935a] J. B. Rosser. A mathematical logic without variables, Part I. *Annals of Mathematics*, 36:127–150, 1935.

[Rosser, 1935b] J. B. Rosser. A mathematical logic without variables, Part II. *Duke Mathematical Journal*, 1:328–355, 1935.

[Rosser, 1936] J. B. Rosser. Extensions of some theorems of Gödel and Church. *Journal of Symbolic Logic*, 1(3):87–91, September 1936.

[Rosser, 1942] J. B. Rosser. New sets of postulates for combinatory logics. *Journal of Symbolic Logic*, 7:18–27, 1942.

[Rosser, 1984] J. B. Rosser. Highlights of the history of the lambda-calculus. *Annals of the History of Computing*, 6:337–339, 1984.

[Sanchis, 1963] L. E. Sanchis. *Normal combinations and the theory of types.* PhD thesis, The Pennsylvania State University, 1963.

[Sanchis, 1964] L. E. Sanchis. Types in combinatory logic. *Notre Dame Journal of Formal Logic*, 5:161–180, 1964.

[Schönfinkel, 1924] M. Schönfinkel. Über die Bausteine der mathematischen Logik. *Mathematische Annalen*, 92:305–316, 1924. English translation, "On the building blocks of mathematical logic," in Jean van Heijenoort, editor, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, pages 355–366, Harvard University Press, Cambridge, MA and London, 1967.

[Seldin, 1968] J. P. Seldin. *Studies in Illative Combinatory Logic.* PhD thesis, University of Amsterdam, 1968.

[Seldin, 1975a] J. P. Seldin. Arithmetic as a study of formal systems. *Notre Dame Journal of Formal Logic*, 16:449–464, 1975.

[Seldin, 1975b] J. P. Seldin. Equality in $\mathcal{F}_{22}$. In H. E. Rose and J. C. Shepherdson, editors, *Logic Colloquium '73*, pages 433–444. North-Holland, 1975.

[Seldin, 1977] J. P. Seldin. The Q-consistency of $\mathcal{F}_{22}$. *Notre Dame Journal of Formal Logic*, 18:117–127, 1977.

[Seldin, 1980] J. P. Seldin. Curry's program. In J. R. Hindley and J. P. Seldin, editors, *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, pages 3–33. Academic Press, London, 1980.

[Seldin, 2000] J. P. Seldin. On the role of implication in formal logic. *Journal of Symbolic Logic*, 65(3):1076–1114, 2000.

[Seldin, 2003] J. P. Seldin. Curry's anticipation of the types used in programming languages. In *Proceedings of the Canadian Society for the History and Philosophy of Mathematics, Volume 15, Twenty-Eighth Annual Meeting, University of Toronto, Toronto, Ontario, 24–26 May 2002*, pages 148–163. Canadian Society for the History and Philosophy of Mathematics, 2003.

[Shapiro, 2000] Stewart Shapiro. *Thinking About Mathematics.* Oxford University Press, Oxford, England, 2000.

[Smullyan, 1961] R. Smullyan. *Theory of Formal Systems.* Princeton University Press, Princeton, New Jersey, 1961.

[Stegmüller, 1964] W. Stegmüller. Remarks on the completeness of logical systems relative to the validity-concepts of P. Lorenzen and K. Lorenz. *Notre Dame Journal of Formal Logic*, 5:81–112, 1964.

[Titgemeyer, 1961] R. Titgemeyer. Über einen Widerspruch in Cogans Darstellung der Mengenlehre. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 7:161–163, 1961.

[Troelstra and van Dalen, 1988] A. S. Troelstra and D. van Dalen. *Constructivism in mathematics: An introduction*, volume I. North-Holland, Amsterdam, 1988.

[Turing, 1936–1937] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42:230–265, 1936–1937. Correction, *Ibid.* 43:544–546, 1937.

[Turing, 1937a] A. M. Turing. Computability and λ-definability. *Journal of Symbolic Logic*, 2(4):153–163, December 1937.

[Turing, 1937b] A. M. Turing. The $\mathfrak{p}$-function in λ-K-conversion. *Journal of Symbolic Logic*, 2(4):164, December 1937.

[von Neumann, 1925] J. von Neumann. Eine Axiomatizierung der Mengenlehre. *Journal für die Reine und Angewandte Mathematik*, 154:219–240, 1925. Correction, *ibid.* 155:128. English translation: An axiomatization of set theory, in J. van Heijenoort, editor, *From Frege to Gödel: A Source Book in Mathematical Logic 1879–1931*, Harvard University Press, Cambridge, Mass., USA, 1967, pp. 393–413.

[von Neumann, 1928] J. von Neumann. Die Axiomatizierung der Mengenlehre. *Mathematische Zeitschrift*, 27:669–752, 1928. Based on the author's doctoral dissertation in Hungarian, University of Budapest, September 1925 (author's footnote, p. 669). Reprinted in A. H. Taub, editor, *The Collected Works of J. von Neumann*, volume 1, Pergamon Press, Oxford, England, 1961, pp. 339–422.

[Whitehead and Russell, 1910–1913] A. N. Whitehead and B. Russell. *Principia Mathematica*. Cambridge University Press, Cambridge, England, 1910–1913. Second edition, 1925–1927.

PARADOXES, SELF-REFERENCE AND TRUTH IN THE 20TH CENTURY

Andrea Cantini

This is the reason why the notion of truth is so precious to us: it is one of the means by which finite minds are able to apprehend the infinite. (W. O. Quine, ***Philosophy of Logic***)

Here is a crude list, possibly summarizing the role of paradoxes within the framework of mathematical logic:

1. directly motivating important theories (e.g. type theory, axiomatic set theory, combinatory logic);
2. suggesting methods of proving fundamental metamathematical results (fixed point theorems, incompleteness, undecidability, undefinability);
3. applying inductive definability and generalized recursion;
4. introducing new semantical methods (e.g. revision theory, semi-inductive definitions, which require non-trivial set theoretic results);
5. (partly) enhancing new axioms in set theory: the case of anti-foundation AFA and the mathematics of circular phenomena;
6. suggesting the investigation of non-classical logical systems, from contraction-free and many-valued logics to systems with generalized quantifiers;
7. suggesting frameworks with flexible typing for the foundations of Mathematics and Computer Science;
8. applying forms of self-referential truth and type-free comprehension in Artificial Intelligence, Theoretical Linguistics, etc.

Below we attempt to shed some light on the genesis of the issues 1–8 through the history of the paradoxes in the twentieth century, with a special emphasis on semantical aspects.

Handbook of the History of Logic. Volume 5. Logic from Russell to Church
Dov M. Gabbay and John Woods (Editors)
© 2009 Elsevier B.V. All rights reserved.

1 INTRODUCTION

The general aim of the chapter is to describe the emergence of non-trivial logical themes from the discussion on paradoxes, with a special emphasis on the semantical paradoxes and in connection with the foundations of property theory and set theory.

The present historical analysis mainly deals with logical works, which focus on three interacting key-notions: truth, self-reference, type free comprehension. The basic questions we try to answer are:

1. how does the question of paradoxes relate to the main stream of mathematical logic (at different stages of development)?

2. Which relevant logical tools emerge and/or are refined, in strict connection with the discussion of paradoxes about truth and predicate application?

As a final outcome, we would like to reach a reasonably complete catalogue of ideas, methods and logical tools, involved in the investigation of paradoxes which took place in the past century. It should also come into view that the role of semantical paradoxes for a modern mathematical development of definability theory (in its various form: recursion theory, descriptive set theory) is certainly not negligible and it arises in close correlation with set theory, model theory, recursion theory.

Since the literature on (property theoretic and semantical) paradoxes, even if restricted to the past century, is overwhelming, we have to make a twofold choice: 1) we concentrate upon contributions, which have a prominent technical, systematic vein; 2) we mainly consider works which lie the range of Mathematical Logic, in the standard sense, thus including Set Theory, Proof Theory, Recursion Theory and Model Theory. This restriction has to be made explicit because of a considerable shift of emphasis in the 20th century logical research. While the issue of paradoxes comes to the scene mainly in the course of the foundational debate at the beginning of the 20th century, the problem has been considerably transmuted, especially during the final quarter 1975-2000, by the interaction with Computer Science, Artificial Intelligence and related disciplines

A further methodological remark is in order. When dealing with paradoxes, there is an obvious danger: one can emphasize the dramatic consequences and hence overestimate their role in the development of mathematical logic, or else assign them a place in the corner of logical games and puzzles, at best an heuristic role. The truth, as is often the case, lies in the middle, and historical analysis is of invaluable help to decouple the two different attitudes.

Another preliminary observation concerns the problem of *forerunners*. If we look at the history of Logic, we may soon arrive at the conclusion that all the basic solutions of, say, the Liar paradox, are *there* (i.e. there is hardly anything new to say, after a(n about) 2500 years long logical debate); hence we might be deceived by a continuity feeling in the traditional analysis of paradoxes. *But* an essentially new point is that Mathematical Logic has produced *mathematical*

models thereof, and not simply a logico-linguistic or conceptual analysis. Models introduce artifacts, and, as usual in formal modelling, there will be a question of adequacy; on the other hand, models act as microscopes: novel distinctions and problems arise. Above all, precise statements and results are born out: *the characteristic shift from antinomies to theorems of negative nature (impossibility results)*.

With this in mind, we conclude by anticipating that there is a fundamental difference between the role of paradoxes before, say, 1897-1903 (dates of the first published paradoxes, the Burali-Forti paradox, and of Russell's paradox) and the subsequent discussion: for the first time the analysis of simple paradoxes, usually regarded as idle or scholastic, touches the living flesh of a mathematical theory *in fieri* and, soon, Russell will officially introduce the issue of "contradictions" ([Russell, 1908]), as a challenge for the foundations of mathematics and logic. Moreover, it happens that paradox-like arguments will later find their own place as (tricky) proof methods in concrete areas of mathematical logic, in particular in recursion theory.

WARNING 1. No claim of completeness is made. In particular, the following items, though obviously linked to the keywords of this chapter, have been omitted:

- recent developments of (illative and non-illative) combinatory logic;
- the treatment of self-reference and provability, and hence of provability logic in its algebraic and modal versions;
- paradoxes of vagueness (Sorites, Heap) as well as paradoxes of decision and choice (the prisoner's dilemma, Newcomb's paradox; cf. [Sainsbury, 1988]);
- nothing is said about set theoretic paradoxes in the sense of standard set theory and its extensions; the same holds also of classical positive consequences of ZFC, usually termed as paradoxes, e.g. the Banach-Tarski paradox (cf. [Wagon, 1985]);

WARNING 2. During the last fifteen years, substantial work is being carried out in connection with the history of logic in the first half of last century (especially on Hilbert's school, Gödel, Tarski, etc.) on the basis of unpublished material. Since this kind of research is ongoing and a large part of the present work has been written before 1995, we have mostly limited our analysis to *published sources*, giving however references, whenever possible, to recent research papers based on unpublished material.

2 EARLY DEVELOPMENTS (1897-1930)

Traditionally, paradoxes can be classified according to the so-called Peano-Ramsey distinction [Peano, 1902-1906; Ramsey, 1926]) between semantical paradoxes,

which ought to belong to linguistics/epistemology, and logical (set-theoretic) paradoxes, which ought to belong to mathematics in the proper sense. However, from an historical perspective, it is much more appropriate to partition them according to the motivating problems; furthermore, it will appear that, contrary to what is generally stressed, also semantical paradoxes are linked with vital mathematical notions (the continuum above all).

We anticipate that this period is characterized by the fact that local mathematical difficulties (concerning the notion of order type, cardinal number, the arithmetized continuum) quickly merge into antinomies of a general logical import. It is not exaggerated to ascribe to Russell the responsibility of this change. The historical analysis shows that the antinomies first appear in the context of refining Cantorian set theory and induced specific technical problems. Then almost immediately, the unintended by-product is that Frege's theory is deathly wounded, Cantorian set theory shows an intrinsic mathematical richness to survive on and to be axiomatized, while, at the same time, fairly innovative logical issues, like self-application and definability, emerge as crucial for the research to come.

After the first, say, eight years of the century, paradoxes stably become a kind of standard tool, that leads to distinguish between consistent vs. inconsistent totalities and to the enucleation of suitable disciplines for the comprehension principles (limitation-of-size, no-classes solution, type theories). Roughly speaking we shall discuss paradoxes pertaining to the following notions:

1. ordinal and cardinal numbers (Burali-Forti [1897b], Cantor 1895/6, 1899, see [1932;1962], pp. 443-447, 451);

2. definability and the arithmetic (or atomistic) continuum (Richard [1905], König [1905b], [1905a], [1906], Bernstein [1905c], Berry [1906b], Grelling and Nelson [1908]);

3. the logical notions of property, set, class (Russell 1901 [1993], Zermelo [1902]; Husserl [1979, p. 399]; [Frege, 1984^2, p. 80]);

4. the notions of proposition and truth (Russell [1903], [1906b], [1908])

2.1 1897-1903: the emergence of the contradictions

The first published set theoretic antinomy concerns the class of the ordinal numbers and is due to Cesare Burali-Forti,[1] a mathematician and logician of the Peano school. The background is the ongoing project of formalization of (even advanced) mathematical practice in Peano's logical symbolism (in the special case, a logical analysis of the notion of order types and ordinal). Curiously enough, Burali-Forti was not explicitly aiming at a paradox; instead, he intended to prove that there are ordinal numbers which are incomparable with respect to their natural ordering: so that the class of all ordinal numbers (and in general the class of all ordered

[1] C.f. [Moore and Garciadiego, 1981].

types) could not be linearly ordered. His argument runs by contradiction and, in standard terminology, is readily rephrased as follows: assuming that the class ON of all ordinals could be linearly ordered, ON itself would be well-ordered and it would possess an ordinal $\Omega \in ON$. Thus ON would be similar (order-isomorphic) to a proper initial segment of itself, the one determined by Ω, contradicting a well-known theorem about well-ordered sets.

Two remarks[2] are in order here. First of all, the notion of "well-ordered set" (and "ordinal number") was not correctly defined by Burali-Forti: he just missed a crucial second order condition, implying that in a well-ordering $(W, <)$ every non-empty $X \subseteq W$ has a first element (in the sense of $<$).[3] On the other hand, Burali-Forti quite soon acknowledged his omission (in a subsequent note of the same year [1897a]), and stated without proof that his results were still sound for properly well-ordered sets.

The second observation is that Burali-Forti's argument raised a genuine difficulty for the notion of transfinite number and for Cantorian set theory in general, even if, as a matter of fact, the proof was ill-based. This is perhaps due to the fact that Burali-Forti's antinomy has a specific mathematical nature: it is not only grounded upon the logical mechanism underlying set theory, but it makes an essential use of basic properties of well-orderings and is concerned with a collection (category) of structures with their own structure preserving maps (morphisms).

Also, looking at the history of the foundational and logical ideas of the period, it is clear that Burali-Forti's paper was not an isolated intuition. Times were ripe for the open discussion of riddles about the foundations of set theory; Cantor's correspondence [1932;1962] (published in 1932 by Zermelo) yields evidence that the difficulty was seriously and independently considered by the father of set theory, as witnessed by his letters to Dedekind (1899). According to [Bernstein, 1905b, p. 187], Cantor had the antinomy as early as 1895, and he would have informed Hilbert in 1896. This is actually witnessed by the correspondence of Cantor ([Cantor, 1991]), where he sketches the argument (letter to Hilbert of 26.9.1897) that the totality of all alephs cannot be at the same time well-defined and conceivable as a completed set (*fertige Menge*).

In a letter to Dedekind of July 28, 1899, Cantor [1932;1962], on the basis of the theorem that ordinal numbers are linearly ordered, stated that the multiplicity of ordinal numbers is itself well-ordered, but he drew the consequence that ON is not a set: hence that no ordinal can be assigned to it.

In a second letter to Dedekind of August 31, 1899 Cantor went further to discuss another set theoretical difficulty, which directly involved the notion of the cardinal number and implied that one cannot consistently think of the "the set of all conceivable sets".

[2] [van Heijenoort, 1967], p. 104.

[3] See Cantor's *Grundlagen einer allgemeinen Mannigfaltigkeitslehre* of 1883, the 1897 paper on order types, that was to appear a bit later than Burali-Forti's work in *Mathematische Annalen*, 46, 481-512, *Beiträge zur Begründung der transfiniten Mengenlehre* and also Cantor's letter to Young, cited by van Dalen-Monna [1972], p. 23, with a severe judgement on Burali–Forti's work.

If the collection S of all cardinals were a set, we could form the set $T = \bigcup\{M_a | a \in S\}$ (M_a being a set of cardinal a). Then T would be *the set of all sets* and would possess a cardinal a_0. But by Cantor's theorem there is a cardinal $b > a_0$. This implies that T would contain a subset M_b of cardinality strictly greater than T's cardinality: contradiction![4]

As is well-known, Cantor made a general point out of his paradoxes, namely that it is necessary to distinguish two kinds of multiplicities ('Vielheiten'): the consistent ones, i.e. sets in the proper sense, and the inconsistent ones. The inconsistent multiplicities are characterized by the fact that they cannot be regarded as one thing ('Einheit'), without falling into contradiction, as in the case of the totality of ordinal numbers, or the totality of all alephs.[5] Of course, Cantor's ideas adumbrate the set/class distinction and are, in a sense, a first attempt to deal with the set theoretic paradoxes and are explicitly mentioned by Hilbert in his 1904 Heidelberg lecture, ([van Heijenoort, 1967], p. 131) at the International Congress of Mathematicians. Hilbert briefly exposes Cantor's proposal, but he criticizes it as largely subjective, since no criterion for the distinction is given (he has in mind an axiomatic treatment, cf. below).

With the second published antinomy (Russell [1903], paragraphs 78, 101–107; Frege [1962], Appendix, dated October 1902) we are driven from Cantor's paradise into the realm of the foundations of logic and the philosophy of mathematics. On June 16, 1902 Russell presented the contradiction in a letter to Frege:

> "Let w be the predicate: to be a predicate that cannot be predicated of itself. Can w be predicated of itself? From each answer, the opposite follows. Likewise, there is no class (as a totality) of those classes which, each taken as a totality, do not belong to themselves. From this I conclude that under certain circumstances a definable collection does not form a totality. " ([van Heijenoort, 1967], p. 124–125)

There are fundamental features in Russell's arguments that are new with respect to Burali-Forti and Cantor, and bring out quite clearly the main ingredients of the contradiction: its relation to the basic mechanism of *predicate application* and its explicit *self-referential* (reflexive, in Russell's future own terms) character. Moreover, Russell's argument is striking, because it is general and simple, and it apparently does not refer to the sophisticated mathematical universe of Cantor and Burali-Forti; so how did Russell get there?

The problem of the origins of the paradox sheds light on the context of a discussion, which initially focussed on the problem of the foundations of set theory, but was soon to shift, under Russell's impulse, towards general fundamental logical themes. Though a detailed reconstruction is beyond the scope of this section (cf. [Grattan-Guinness, 1978], [Moore, 1995] and the most recent [Griffin, 2004]), it is worth summarizing it.

[4]This antinomy is independently rediscovered by Russell in *The Principles of Mathematics* [1903], paragraphs 344-349.

[5]The distinction is also in Schröder [1890], p. 213; but it not quite clear to the present author whether the two notions indeed coincide.

Russell himself states that he got his contradiction by considering the antinomy arising with Cantor's theorem (see [Russell, 1903], footnote 7, par.344; par.100, p. 101). Indeed, it turned out that he was at first entangled in the study of "the contradictions in the relation of continuous quantity to number and the continuum",[6] when he was asked by the French philosopher Louis Couturat to review the dissertation on atomism by the French neo-kantian philosopher Hannequin. At that time Russell learned of Cantor's work, and for a while he rejected the very idea of a mathematical (atomistic) continuum and of transfinite number as self-contradictory. In particular, he was concerned with "the antinomy of infinite number", which follows if one applies the notion of totality to numbers. This antinomy bears a clear resemblance to Cantor's paradox, though[7] Russell does not adopt the notion of cardinal number, nor does he hinge upon Cantor's theorem (using a vague successor notion, instead of cardinal exponentiation): if we could conceive the totality of all numbers and if this totality had a number, such a number ought to be the greatest, say N; but then $N+1$ is greater than N, contradiction.[8] Between November 1900 and January 1901, after the Paris Congress and under the influence of Peano, Russell changes his attitude towards the idea of infinite number and he obtains the paradox of the largest cardinal (see paragraphs 346-349); but he still holds for a while that the class CL of all classes and Cantor's theorem can live together: CL exists and it has a number, but one cannot prove that the number of CL is less than the number of its subclasses, i.e. the theorem does not hold for CL. The final step is accomplished, during the writing of the final draft of *The Principles of Mathematics* (May 1901): here one finds the first version of the paradox.[9]

Yet, as the reader will observe, Russell's contradiction lies dormant for one year. There is evidence in Russell's correspondence that he did not realize the importance of the discovery, until he got the answer from Frege; there Frege stressed the dramatic effects of Russell's observation and he promised an appendix where Russell's discovery is taken into account. Indeed, Frege's second volume of the *Grundgesetze* contains in the celebrated *Nachwort* another presentation of the antinomy and the first attempt towards a solution: he proposes to replace the crucial postulate V with a variant that should be capable of blocking self-application and in the end the contradiction. Frege's idea is that, if two classes defined by the concepts F and G are equal, then $F(a) \leftrightarrow G(a)$, for every object a which is distinct from $\{x|Fx)\}$ and $\{x|G(x)\}$. As shown by Lesniewski in 1938, Frege's variant was doomed to inconsistency (cf. Sobocinski [1949]).

Russell's antinomy, like Burali-Forti's, was independently known in Göttingen

[6] From an unpublished paper, cited in [Moore, 1995], p. 219.

[7] This is already remarked by [Moore, 1995].

[8] [Moore, 1995, p. 226] citing from unpublished Russell's work

[9] Anellis [1991] argues that a first version of the Russell paradox is already in a letter to Couturat, dated 8.12.1900, where Russell mentions to have found a mistake in Cantor and he holds against Cantor: there is a largest cardinal number. Russell notes that Cantor's theorem does not apply if we consider the class of all classes ([Anellis, 1991], p. 39).

since 1902. Hilbert states in a letter to Frege of 7.11.1903 [10] that Zermelo, under his suggestion, had already discovered the same contradiction and a similar remark was made by Zermelo in a footnote contained in his second paper on the well-ordering theorem [1908a]. Most convincingly, there is an indirect confirmation of this fact in a note of Husserl [1979], of 16.4.1902, where Husserl registers the following result of Zermelo (in the context of the discussion of Schröder's work): a set M that comprises as elements all of his subsets is inconsistent. The proof is standard; consider the set M_0 of all elements of M which are not elements of themselves (e.g. the empty set is in M_0). This set is a subset of M and hence by assumption on M, $M_0 \in M$. If $M_0 \in M_0$, then M_0 is not a member of itself. Hence M_0 is not in M_0 and since $M_0 \in M$, $M_0 \in M_0$: contradiction.

Furthermore, as it can be seen from unpublished work of Hilbert (now published and translated in [Peckhaus and Kahle, 2002]). Hilbert not only outlines Zermelo's version of the Russell paradox in the *Vorlesung* of the summer term 1905, but he also presents a new contradiction of his own, together with a remarkable version of the *functional version of the Russell's paradox*, later to become popular in the context of combinatory logic, lambda calculus and recursion theory.[11]

Let us only sketch Hilbert's paradox in modern terms. Let $\mathcal{C}$ be the least collection which satisfies the following closure conditions: (i) the set $\mathbb{N}$ of natural numbers is an element of $\mathcal{C}$; (ii) $X^X \in \mathcal{C}$, whenever $X \in \mathcal{C}$ (X^X being the set of all functions from X to X); (iii) $\bigcup X \in \mathcal{C}$, whenever $X \subseteq \mathcal{C}$. Then trivially $\mathcal{C} \subseteq \mathcal{C}$, whence by (ii), $\bigcup \mathcal{C} = \mathcal{U} \in \mathcal{C}$ and finally $\mathcal{F} = \mathcal{U}^{\mathcal{U}} \in \mathcal{C}$. But by definition of union, $\mathcal{F} \subseteq \mathcal{U}$; hence there would exist a map of $\mathcal{U}$ onto $\mathcal{F}$ and a contradiction could be easily derived by diagonalization à la Cantor.

According to Hilbert, the paradox is more significant and convincing than other contradictions, just because it is *purely mathematical,* i.e. it involves only mathematical objects and mathematical operations. He also adds that, after believing that it would constitute a stumbling block for set theory, he is confident in a solution after a revision of the foundations. Today, in the light of Zermelo's axiomatization, we can readily locate the source of the paradox in the step where the union operation is applied to the collection $\mathcal{U}$, which has to be a set: hence

[10] There he thanks Frege for the second volume of the *Grundgesetze der Arithmetik*, see [Frege, 1976, p. 80].

[11] The text of the paradox is published in [Peckhaus and Kahle, 2002]; but see also [Peckhaus, 2004]. A detailed discussion of Hilbert's view on paradoxes, also in the light of Hilbert's course lectures between 1905 and 1920, is given by [Kahle, 2004]. For a reformulation of the paradox using power set operation instead of exponentiation, see [Kanamori, 2004]. As far as we know, Kahle is the first to observe the interesting occurrence of the functional paradox in the version of the 1905 Vorlesung, due to E. Hellinger. The functional paradox is obtained by first assuming that the universe includes everything, i.e. variables range on the same par over objects and functions. Then one introduces a new operation (universal application in our sense): xy stands for the object consisting of the combination of x and y. Having fixed two distinct objects 0 and 1, one makes the crucial assumption that the universe is closed under arbitrary definition by cases and hence there exists an object f such that $fx = 0$ if $xx \neq 0$, and $fx = 1$ if $xx = 0$. Then one chooses $x = f$ (as x ranges everywhere) and easily derives a contradiction. The argument is regarded by Hilbert as a model for clarifying the source of all contradictions; it is worth stressing that it is based on functional self-application and hence direct self-reference.

the argument shows that this is not a set.

So why did people in Göttingen refrain from publishing the contradiction? A first ground was probably Hilbert's attitude that contradictions and paradoxes are natural phenomena for a newly born and growing discipline, nothing dramatic in a sense ([Peckhaus, 1990], p. 51). Paradoxes are symptoms of growth, more than a deep malaise; they are thought of as minor, temporary facts, that might only obscure what Hilbert considered to be Cantor's paradise.

More deeply, Hilbert's diagnosis is that the contradictions mainly arise because traditional logic is insufficient and the theory of concept-formation needs to be sharpened. Traditionally, a concept C is given insofar as it is determined of an arbitrary object whether it falls under C or it does not; but according to Hilbert, this is methodologically wrong: *a concept is really fixed in a net with other concepts* (letter of 7.11.1903, in [Frege, 1976], pp. 79-80), and this net is specified by the axioms; so only the consistency of the axioms that define the concept grants the legitimacy of C. The fault is logical, not mathematical. But eventually this requires the development of a grand new research program, his *Beweistheorie*.

Be that as it may, the Russellian contribution of [Russell, 1903] should not be underestimated. The new antinomy - as well as Burali-Forti - is repeatedly discussed in various forms in the *Principles* (sec.78, sec.84-85, sec.101). In sec.85, the contradiction is adapted to show that a propositional function ϕ cannot be regarded by itself as a logical subject (i.e. as separated from its argument; it is not *saturated* in Frege's terms); otherwise, ϕ would apply to itself, $\neg\phi(\phi)$ would be a propositional function and it would be possible to reproduce the inconsistency. It is also interesting to recall that, in the previous section 84, incidentally, Russell observes that, when we speak of the $x's$ such that $\varphi(x)$, we are introducing in the calculus of propositions *the notion of truth*; so in present-day terms, the roots of the difficulties are perceived as semantical as well. In ch.X, section 102, Russell spells out a purely logical form of Cantor's theorem, which captures the logical essence of diagonalization: no binary relation can parameterize all unary predicates (i.e. no binary relation R exists such that for all unary predicates P, there is an object a for which one has: for all x, $R(a, x) \leftrightarrow P(x)$). For then we might argue by diagonalization and construct via R a new unary predicate $D(x) := \neg R(x, x)$ which is not parameterized by R. A variant of this result (also in par.102) immediately applies to propositional functions: "any method by which we attempt to establish a one-to-one or many-one correlation of all terms and all propositional functions must omit at least one propositional function" (ibid. p. 103).

Russell's conclusion points to the critical role of apparently safe logical principles: either one has to give up "the assumption that any propositional function containing only one variable is equivalent to asserting membership for a class defined by the propositional function (i.e. the comprehension principle, that was explicitly stated by Peano); or one has to reject the idea that every class can be taken as one term " (pp. 102-103).

Also, it is utterly clear that in Russell's hands the paradox spreads out and

becomes extremely flexible: it applies to predicates, classes, propositional functions and it forces to delineate a new structure for the logical-mathematical universe, which is outlined in a first exposition of the doctrine of types, as a means to solve the contradiction: "it is necessary to distinguish various types of objects, namely terms, classes of terms, classes of classes, classes of couple of terms, and so on; and that a propositional function ϕx in general requires, if it has to have any meaning, that x should belong to some one type. Thus $x \in x$ was held to be meaningless, because $\in$ requires that the relatum should be a class composed of objects which are of the type of the referent" (p. 107, par.106).

The doctrine of types is put forward tentatively in the second appendix of the *Principles* (par.500). It is assumed that to each propositional function ϕ is associated a range of significance, i.e. a class of objects to which the given ϕ applies in order to produce a proposition; moreover, precisely the ranges of significance form types. However there are objects that are not ranges of significance; these are just the individuals and they form the lowest type. The next type consists of classes or ranges of individuals; then one has classes of classes of objects of the lowest type, and so on.

Once the hierarchy is accepted, new difficulties still arise; in particular, if one accepts that propositions form a type (as they are the only objects of which it can be meaningfully asserted that they are true or false). This is a crucial point and leads Russell to another contradiction, this time explicitly dealing with semantical notions. First of all, there are apparently as many propositions as objects (just consider the map $x \mapsto (x = x)$; p. 527). On the other hand, if it is possible to form types of propositions, as it seems the case, there are more types of propositions than propositions, by Cantor's argument. But we can inject types of propositions into propositions by appealing to the notion of logical product and this leads to troubles. Indeed, let m be a class [12] of propositions and let Πm be the proposition "every proposition of m is true" (which has to be thought of as a possibly infinitary conjunction or logical product). Clearly, if m and n are different classes, the propositions Πm and Πn are different, i.e. the map $m \mapsto \Pi m$ is injective. Then, if we consider the class $\{p | \exists m(\Pi m = p \wedge p \notin m)\} = R$, we have, by injectivity:

$$\Pi R \in R \Leftrightarrow \Pi R \notin R,$$

whence a contradiction.

In order to avoid it, Russell questions the identity of equivalent propositional functions and the nature of the logical product of two propositions. He notices that, if we consider the new class of propositions $m' = m \cup \{\Pi m\}$, then

$$\Pi m' \Leftrightarrow \Pi m \wedge \Pi m \Leftrightarrow \Pi m.$$

So, if one were to adopt the extensional point of view, and hence equivalent propositions should be identified, then $\Pi m = \Pi m'$. As $m \neq m'$, $m \mapsto \Pi m$ would not be injective and the contradiction above could not derived. However, Russell sticks

[12] 'Class' is here the same as 'range' or 'type'.

to an intensional point of view, stressing that equivalent propositions can be often quite different. So it seems that one is forced to reject the assumption that propositions form one type, and hence that they ought to have various types, while logical products ought to have propositions of only one type as factors.

This will be eventually the base of the 1908 solution, but here Russell reputes the suggestion as harsh and artificial; as the footnote of p.527 shows, he still believes that the set of all propositions is a counterexample to Cantor's theorem!

It is thus fair to say that the *Principles of Mathematics*, as its coeval Fregean second volume of the *Grundgesetze* end up with an unsolved antinomy: Russell concludes by declaring that "what the complete solution of the difficulty may be, I have not succeeded in discovering; but as it affects the very foundations of reasoning , I earnestly commend the study of it to the attention of all students of logic." (p.528)

2.2 1903-1905: definability and the continuum

What is the reaction of the mathematical world to the contradictions?

At the Heidelberg International Congress of Mathematicians, Hilbert pleads the axiomatic method as a privileged tool to face the problem of a simultaneous consistent foundation of logical and mathematical notions, and he critically surveys the point of view of several leading scholars, among them Frege, Dedekind and Cantor as directly concerned with the set-theoretic paradoxes. But *he does not explicitly mention Burali-Forti nor Russell*. He limits himself to note that Frege's work is based on the dangerous idea that "a concept (set) is defined and usable if and only if it is determined whether the object is subsumed under the concept or not, and here he imposes no restriction to 'every' ..." ([Hilbert, 1905], [van Heijenoort, 1967], p.130). Thus Frege must assume the existence of the set of all sets, and he entangles himself with the set-theoretic paradoxes. Hilbert's conclusion is that "a major goal of the investigations into the notion of number should be to avoid such contradictions and to clarify these paradoxes"(ibid., p.130). He then proceeds by pointing out that Dedekind relies upon insecure grounds, as he employs the notion of the totality of all objects, while Cantor sensed the contradictions (as we have already recalled), but he proposed a distinctions between consistent and inconsistent totalities, which is not mathematically acceptable yet. On his side, Hilbert then sketches the idea of combinatory consistency proofs as means to justify axiomatized theory and hence mathematical existence (see also [Sieg, 1999]).

The involvement with foundational puzzles was destined to increase during the Congress and in its immediate temporal neighborhoods; furthermore, if the new contradictions touched the notions of ordinal and cardinal number, and they embrace the very conception of set and genuinely logical concepts, the difficulties soon propagated, up to include the notion of *definability*, and its relation with a fundamental issue: the structure of the mathematical continuum and Cantor's continuum hypothesis.

At the Heidelberg Congress Julius König claimed to have disproved the existence of a well-ordering of the continuum, thus refuting Cantor's hypothesis. Due to a mistake discovered by Zermelo, his paper was immediately withdrawn, but the subsequent year, König [1905a] produced a new proof, that was presented in June 20, 1905 to the Hungarian Academy of Sciences.

His argument essentially goes as follows. Consider the reals which are definable in finitely many words. They clearly form a countable set and they can be enumerated in a sequence: $E_0, \ldots, E_n, \ldots$. Since the continuum is uncountable, there exist reals not occurring in the given enumeration. Assuming that the continuum is well-ordered, there exists "the least real E which is not in the sequence$\{E_n\}_{n\in\omega}$"; but the very expression "the least real E which is not in the sequence$\{E_n\}_{n\in\omega}$" defines E with finitely many words; so E occurs somewhere in the sequence: contradiction !

On the other hand, König observes that the argument extends to the second number class (or, in general, to any uncountable well-ordered set). The collection FOD of finitely definable countable ordinals is certainly denumerable; hence there would exist the least ordinal α such that $\alpha \notin FOD$. But then α would be finitely definable and an element of FOD: contradiction! In this case, taking inspiration from Cantor, König's diagnosis is that one ought to distinguish between totalities that can be conceived as completed and those that are never understood as finished. But he is apparently suggesting something more ([König, 1905a], p. 148): a completed totality or set in the proper sense should provide not only a rule for defining its elements, but also a means for distinguishing them.[13] According to him, Cantor's second number class is probably not a set, because we only have a "collective notion", which allows to distinguish between elements and non-elements of the class.

Less than two months later Zermelo had conceived his (first) proof of the well-ordering theorem from the axiom of choice.[14] So there was an evident conflict ! Clearly, since the well-ordering theorem is correct, the conclusion of the Hungarian mathematician must be false; yet, it still remains a contradiction in which the notion of definability with finitely many words plays a major role.

A contradiction, which is strictly related to König's had been published a bit earlier by Jules Richard, a mathematics teacher at the Lycèe of Dijon.[15] He observes – using an enumeration of all permutations with repetitions of the twenty-six letters of the French alphabet – that the set E of reals that can be defined by finitely many French words is denumerable and hence we can assume to have an enumeration $u_1, u_2, \ldots$ of all those numbers.

But now a contradiction easily arises by forming the following real N: the integer part of N is 0, while its n-th decimal digit of N is $p + 1$ if u_n has the nth decimal digit p different from 8, 9; else, the nth-decimal digit of N is 1. By

[13] This idea is also shared by Poincaré.

[14] Not yet stated as such; cf. [Levi, 1902] and four years later [Zermelo, 1908a].

[15] [Richard, 1905] was communicated in the form of a letter to the *Revue Générale des Sciences Pures et Appliquées*, dated March 30, 1905.

construction, N will not occur in $u_1, u_2, \ldots$. On the other hand, if we consider that N is defined by means of a finite collection of letters, this must occur in $u_1, u_2, \ldots$

It is to be stressed that Richard does not rely upon the well-ordering of the continuum; and the solution he proposes is interesting for the foundational debate to come. *He points out that the definition of N mentions the very collection E, to which E must belong*; but E is yet to be defined; and the definition of N makes sense, only if the set E is completely defined. This cannot be done except by means of infinitely many words. In conclusion, the contradiction disappears, since the definition of N has no definite meaning.

To make the picture of the foundational debate in the neighborhood of Heidelberg Congress more intricate and complete, one should not forget that, at a given time in 1905, there was not only the contrast between Zermelo and König; indeed, Felix Bernstein published in 1905 a short note [1905a] *Die Theorie der reellen Zahlen*, where Cantor's continuum hypothesis was claimed to be settled in the positive![16] Again, Bernstein's argument crucially relies upon the notion of definability. First of all, he identifies the continuum with the space N^N of all numbers theoretic functions; but he criticizes the so called Dirichlet's notion of arbitrary function and he believes that it is possible to give a foundation to the continuum using only *computable* reals, i.e. reals possessing an explicit *Bildungsgesetz*. According to him, this is not restrictive since – he claims – there are non-denumerably many computable reals in his sense. Then he also states that it is possible to display the new computable continuum in a hierarchy (i.e. a $\subset$-increasing sequence) $\{B_\alpha\}_{\alpha<\omega_1}$ of subsets of N^N, which is defined in such a way that each B_α is at most of cardinality $\aleph_1$ and hence that the cardinality of the union of the sequence is at most $\aleph_1$. The idea is to start with a basic domain B_0 of simple functions $\subseteq N^N$ (e.g. those having finitely many values), then to define a new domain B_1 by closing off under the operations of B_0, together with operations, which are recursively defined from elements of B_0, and so on. Of course, Bernstein's claims are hardly justified, as the notion of definability involved is essentially vague; and his note is after all only a sketch of a proof. A posteriori, one might observe that he got at least *in nuce* a sound intuition: it is possible to settle the continuum problem, if one has an appropriately general notion of definability (or computability, constructibility). Other works dealing with the problem of definability are those of [Hobson, 1905], [Dixon, 1907], that prompted Russell[1906b].

Why definability? On one hand, this is ultimately connected to the issue of a better understanding of the atomistic continuum; the arithmetized conception necessarily shifts the attention towards *arbitrary infinite sequences of natural numbers* (or arbitrary countable sets of natural numbers) as main individual mathematical constituents, underlying the mathematical practice (not simply the Cantorian transfinite).

But this is not so easy to swallow; there remains the idea that an infinite

[16] In [1905b] Bernstein proposes a solution of Burali-Forti's antinomy which resorts to assuming that the ordinal β of the set ON of all ordinals cannot be increased, i.e. $\beta = \beta + 1$; then he tries to argue that the set of all subsets of ON cannot well-ordered.

sequence (or an infinite set) has to have a forming rule. But what is a rule? Since one must be extremely liberal (in order not to have special classes of reals, if one is too restrictive), one is naturally driven to think of arbitrary finitely described laws, shifting the attention towards the syntax of the rules. But no syntax, unless you consider the ordinary language, is available and this engenders indeterminacy (this is Peano's diagnosis [1902-1906]).

On the other hand, the need of a specification of infinite sets appears crucial in the discussion of the related well-ordering problem, nor one should forget the problem of the classification of discontinuous functions and analytically representable functions of real variables, tackled by Baire and Lebesgue in the same years, so deeply intertwined with the question of mathematical existence in general.[17] As Lebesgue remarks (p. 153, [Borel, 1914]), if you only know the defining property of a set, you know nothing of the elements, except what is needed to distinguish them from elements outside the set. Moreover, the question is: "peut on demontrer l'existence d'un être mathématique sans le définir?" After Zermelo's proof, Borel noticed in [1904] that Zermelo only proved the equivalence of the well-ordering problem with the axiom of choice (the possibility of arbitrary choices); he also quoted a letter of Baire, manifesting a skeptical attitude towards Cantor continuum hypothesis: why virtual totalities, determined by completely different notions, ought to have "a common measure"? (p. 516).

2.3 1905-1908: between conceptual analysis and technical tools for solving paradoxes

In the light of the foundational debate, the three years to come are rich of seminal work: new paradoxes arise (Berry, Grelling-Nelson), a unified view of the contradictions of logic and mathematics is masterfully outlined in the opening section of Russell's 1908 paper (see [Russell, 1908] and the preceding [Russell, 1906a]), while a first conceptual distinction between two kinds of paradoxes is set forth in Peano's paper of 1906. Also, the basic ideas of predicativity emerge in the discussion between Poincaré, the leading mathematician of the time, and Russell. Even more important for the history of mathematical logic, there appear fundamental technical contributions for solving paradoxes and shaping the foundations of logic and mathematics: Russell's theory of (ramified) types and Zermelo's axiomatization of set theory. This is not the full picture; refinements and variations of foundational and logical interests are to be found in the works of Borel, Bernstein, Schönflies, Korselt, Levi, Hessenberg, Brouwer and others.

Since we cannot go into a detailed analysis of each single contribution and we deal mainly with paradoxes, we only concentrate on some crucial papers.

[17] See the celebrated letters of 1905 among Baire, Borel, Lebesgue, Hadamard, [Borel *et al.*, 1905] and [Borel, 1914]; Hadamard holds that the existence of mathematical objects is independent upon the means we use to define them; also, according to him, ON is not a set, following Cantor, cf.p.158 [Borel, 1914], *contra* Bernstein. For a historical analysis of the relevant mathematical notions involved, measure and integration, see [van Dalen and Monna, 1972].

Russell and Poincaré on contradictions

The backbone of the discussion is the Poincaré-Russell debate in the *Revue de Métaphysique et Morale*; roughly, it comprises:

1. the long essay *Les mathématiques et la loqique* in the *Revue* of 1905-1906 ([Poincaré, 1905], [Poincaré, 1906a]), where Poincaré strongly criticizes the axiomatic and logicistic foundations (Hilbert, Peano, Couturat, Russell being the major opponents) and vindicates the Kantian point of view about the foundations of mathematics;

2. Russell 1906's paper [1906b] *On some difficulties in the theory of transfinite numbers and aggregates*;[18]

3. the counterattack of Poincaré [1906b], with the subsequent Russellian papers in the *Revue* of 1906, on the *American Journal of Mathematics* (1908), and Poincaré [1910], [1909a], [1912].

While mathematicians mainly focus on the problem of the contradictions, insofar as it touches the foundations of set theory, Russell (1906) proposes a thorough re-examination of the paradoxes, holding the view that his own contradiction "is a special case of a general type of contradictions which result from supposing that certain propositional functions determine a class, when, in fact, they do not do so" ([Russell, 1906b], p. 33). A propositional function may be well-defined for every argument, and yet the collection of the values for which it is defined need not be a class. So the real problem becomes logical: to give a consistent theory of admissible propositional functions.

In this framework, Russell speaks for the first time of the notion of 'predicative norm' ('norm' being the term used by Hobson [1905]) or *predicative propositional function*, as the one which defines a class (hence its negation is also predicative); then the contradictions, like Burali-Forti and his own antinomy, show exactly that certain functions are not predicative.[19]

Russell's diagnosis in [1906b] points to diagonalization and self-reference as major sources of impredicativity: "there are self-reproductive processes and classes; there are properties, such that, given any class of terms all having such a property, we can always define a new term having the property in question. Hence we can

[18] The paper is apparently triggered by a kind of review of Hobson's paper *On the arithmetic continuum* [1905], which is, in turn, motivated by König's 1905 paper [1905a]; cf. also Dixon [1907]

[19] A common generalization of Burali-Forti and his own antinomy is offered: "given a property ϕ and a function f, such that if ϕ belongs to all the members of u, f'u always exists, has the property ϕ and is not a member of u; then the supposition that there is a class w of all terms having the property ϕ and that f'w exist leads to the conclusion that f'w both has and has not the property ϕ " [Russell, 1906b], p. 35 (e.g. choose f ='identity', ϕ=' x is not a member of x', and, also f'u =the ordinal of u, $\phi = x$ is an ordinal", ([Russell, 1906b], p. 35). This is not especially illuminating except for pointing to the essential role played by the existence of the class $\{x|\varphi(x)\}$.

never collect all terms having the said property into a whole "(p. 36). He also observes that, even if a propositional function ϕ is provably impredicative, it might happen that the terms satisfying ϕ could be arranged in a series ordinally similar to a *segment of the series of all ordinals* and hence that it could be non-contradictory to assume that ϕ is in fact predicative. Therefore, even if a propositional function is provably impredicative, it cannot be excluded that *extensionally*, it determines a predicative function. It is clear that if we replace 'it cannot be excluded' by 'it must be the case that' this rather obscure remark could be regarded as a way of postulating a form of the reducibility axiom.[20]

Given that paradoxes are not essentially arithmetical, but logical, Russell tentatively advances three different proposals: the zigzag theory, the theory of limitation of size and the no-classes theory, eventually his own final choice. But the paper does not yield clear cut characterizations. While zigzagness fosters the idea that the right functions ought to be "simple",[21] according to the second view, 'predicative' would be characterized by a certain limitation of size (e.g. the collection of all ordinals ON is too large). In the no-classes theory, classes are not independent entities and anything said about propositional functions is to be regarded as an abbreviation for a statement about its values.

Poincaré [1906b] strongly reacts against Russell, but he proposes *a reinterpretation of predicativity* (as introduced rather neutrally in zigzag theory) as *absence of vicious circles*, thus somewhat extending Richard's diagnosis (cf. above [Richard, 1905]). In essence, he advances a bold generalization, arguing that *all the contradictions* rely upon impredicative (circular) definitions and that the problem with impredicative classes is not that they are empty, but that they have *an undecided frontier* ([Poincaré, 1906b], p. 310). The intuition is not supported by logical analysis, since Poincaré is a strong opponent of formal methods and primarily interested in a philosophical debate. Thus the contradictions offer to Poincaré a means to defend an intuitionistic, Kantian point of view, arguing that the logical definition of natural numbers, as those numbers which belong to all recurrent classes, is impredicative (p. 309), and that number theoretic induction as well the axiom of choice (see. p. 315) constitute independent intuitions, truly synthetic a priori judgements.

It is worth mentioning that at page 305, [1906b], Poincaré assimilates the form given by Russell to the paradoxes (including Russell's, Burali-Forti, König) to that of the Epimenides, i.e. to a semantical traditional puzzle.[22] Thus the challenge of paradoxes tacitly propagates back to the realm of pure, traditional logic, and Poincaré's observations are highly influential, as witnessed by Russell's reply of 1906 (about *Les paradoxes de la Logique*, [Russell, 1906a]).

[20] See below, 2.4.

[21] Gödel (1944) (see [Gödel, 1944]) suggests that this theory ought to be qualified as 'intensional' and regards Quine's system NF of [Quine, 1937] as a candidate for making the idea of zigzagness precise.

[22] Quoting from Russell [1908]: "Epimenides the Cretan said that all Cretans were liars, and all the other statements made by Cretans were certainly lies. Was this a lie?".

Russell [1906a] not only sticks to the view (p. 635) that paradoxes are more on the side of logic than of mathematics, but, typically, rephrases Poincaré's ideas and criticism within the language of formal logic; for instance, the problem of admitting the actual infinity resorts to accept that the word 'all' applies to all or some terms having a certain property, therefore posing the question of a proper understanding of higher order quantification. Moreover, the vicious circle principle ("whatever concerns every or some or any member of a class cannot be a member of the class", p. 634) is stated by explicitly referring to Poincaré, but, at the same time, to the terminology and notions of formal logic in the Peano's sense:

> *whatever comprises an apparent variable should not be one among the possible values of that variable.*

In contrast to the previous paper, Russell definitely adheres to the no-classes theory. He interprets ordinary statements about classes without assuming that classes are independent entities; meaningful propositions concerning classes are regarded as propositions concerning their elements (all of them, some of them). This means that classes are to be eliminated by the corresponding definitions, using as technical tool the 'substitution method' (p. 636), and the contextual elimination of definite descriptions (he cites on p. 637 his paper *On Denoting*, [1905a]). Roughly, if we have a proposition p and a constituent a, Russell (p.637) calls p/a *substitution matrix* and assimilates it to the class, whose members are exactly the values b which, when substituted to a in p, make p true.[23] Nonetheless, class theoretic paradoxes are then solved, insofar as, according to the vicious circle principle, the values must be of a type different from that of the corresponding matrix (class) and classes cannot be members of themselves.

Although Russell endorses Poincaré's view of paradoxes as *Insolubilia*, and he even credits the idea of vicious circularity as a ground for contradiction to Ockham,[24] he does not share the idea that actual infinite is an essential ingredient of the foundational malaise; contradictions arise also if the infinite is not involved and this is clearly shown by the Liar paradox in the form "I lie" (p.632). As far as we know, it is exactly at this place that the most cited semantical puzzle in the history of logic regains a conspicuous position in the logical analysis of the twentieth century.

Here is Russell's analysis of the Liar (p. 642).[25] First of all, he assumes that there exist certain entities that are claimed by genuine statements (e.g. when I say that Socrates is mortal, there is a fact corresponding to my assertion and it is this fact that the proposition claims). He also concedes this, even if the assertion is

[23]This is actually a published trace of the *substitutional theory of classes and relations* that was developed by Russell in unpublished work (see [1905b]) and then dismissed as a perilous route, as it ended up in a new contradiction by a suitable self-referential trick. For details and references, see [Pelham and Urquhart, 1994], [Landini, 2004], [Rouilhan, 2004].

[24]His source is Baldwin's *Dictionary of Philosophy and Psychology* [Baldwin, 1905].

[25]For the sake of historical information, we recall that there is a Russellian manuscript with the appealing title *The Paradox of the Liar*, which the present author has not been able to see, but it is announced as forthcoming in volume V of Russell's *Collected Papers*.

negative; but *not if we deal with existential or universal statements, which contain 'apparent variable'*. In our terminology, *propositions are not closed under quantification*: we must be sure that the collection we quantify upon, is a well-defined totality, as a class. This is not the case when we consider the collection of propositions. Indeed, the Liar can be rendered as 'there is a proposition that I state and that it is false' (p.643). But this contains a quantification (an apparent variable) over the collection of all propositions and, according to the vicious circle principle, we have no right to make such a claim. In other words, the Liar proposition can be false (true) if I make a true (false) statement, or else if I fail in making a statement at all. In the first case, we get a contradiction. It is the second case that Russell accepts: the Liar is false because the statement does not yield a proposition (it fails to be a true claim about the world), as it implicitly contains a quantification over an impredicative collection.

As to Richard's paradox, Russell points out that the class E is not well-defined: the reason is that in the definition of E one employs the very notion of definition, and this is not a well-defined notion (p. 645). Similar considerations apply to the paradox suggested by Berry, which is here stated for the first time in published form, and has the merit not to overcome the domain of finite numbers. Consider the natural numbers that are defined by means of less than 18 syllables; this set is non-empty and finite. So there exists numbers which are not definable using less than 18 syllables. Consider the least such number: clearly, by definition, it is not definable with less than 18 syllables. On the other hand such a number is definable with less than 18 syllables, since it is uniquely determined by the expression "the least number not definable with less than 18 syllables", which has itself less than 18 syllables.

Variations I: Peano, Schönflies, Brouwer, Borel

While Russell and Poincaré analyse contradictions as deeply intertwined with philosophical and foundational ideas (the role of actual infinity, the nature of classes), Peano's analysis of Richard (cf. [Peano, 1902-1906] *Additione a Super Theorema de Cantor-Bernstein*) is an attempt of concretely applying *formal methods* and in particular the rules of the theory of definitions. According to him, Richard's solution of the contradiction is not satisfactory. Since the diagonal Richard number N is defined in the natural language, plus the name for the set E of finitely definable reals in $(0, 1)$, Peano tries to get rid of circularity and reference to E, by fixing a kind of explicit "gödel-numbering": given a natural number n, write n in base B, for B sufficiently big (so to include the number of alphabet letters plus punctuation marks). So each number determines a suitable finite string of symbols in the natural language, and, in certain cases, this string will define a number. Let $Val(n) =$"the decimal number which is defined by the expression encoded by n and interpreted according to the rules of the natural language"; then Richard's number N is defined as the unique real R in $(0, 1)$, satisfying a suitable condition $\varphi(R)$, depending on Val. In order to get the paradox,

one has to show that there exists at least a real satisfying φ and that R is unique. But no such a real can exist: if N were well-defined, and, say, $Val(m) = N$, for some m, the mth-digit c of N would have to be the anti-digit of c itself, i.e. it ought to satisfy the inconsistent $c \neq c$. Hence, Richard's definition is defective in the same sense as "the greatest prime number" is; for N we fail to produce an existence proof. According to Peano, the weak point with Richard's definition is that it is partly formal, partly non-mathematical, hence ambiguous, [26] because it relies on our understanding of natural language: in this sense, one ought to say that Richard's example pertains to Linguistics, not to Mathematics (this view is usually mentioned as inaugurating the distinction between set theoretic or mathematical antinomies, like those of Russell, Cantor, Burali-Forti, and semantical ones, e.g. Richard's, the Liar).

On the general foundational level, Peano adds a criticism to Poincaré's "vrai solution" according to which the problem with the contradiction would lie in the presence of impredicative notion; in fact, he notices that many usual definitions in mathematical texts are impredicative in Poincaré sense (e.g. when I define the least common multiple of two natural numbers a and b as the least number k such a divides k and b divides k, then I refer to the set of all the multiples of a and b to which k belongs). In conclusion, Poincaré's criterion is both too loose (it does not forbid circularity) and too strict (it forbids standard definitions in the mathematical practice).[27]

Schoenflies [1906] moves from the idea that the solution of contradictions must be possible without the axiomatic method. He states that the root of paradoxes lies in the application of logic to contradictory concepts. The main thesis is that it is impossible that there are sets which are members of themselves. When a set is formed, it is 'begrifflich invariant'; it forms a new concept, which adds to those used in forming it. The claim is rather obscure, but it sounds suggestive of an *iterative conception* of set. Of course, once this is accepted, the Russell set of those sets which are not members of themselves *is the universal set* V. But then $V \in V$ belongs to himself: this is impossible and solves Russell's antinomy.

In a subsequent paper Schönflies [1911] strongly criticizes the "philosophical set theorists" (p. 244) Peano and Russell, and Zermelo himself, insofar as he is under their influence. Nevertheless, [Schönflies, 1911], p. 245 proposes a kind of rudimentary axiomatization for set theory, where $a \in a$ is excluded. But he rebuts Russell's views because of their scholastic and unhealthy influence ("unheilvollen Einfluss") on mathematics: the paradoxes do not involve mathematics at all; they derive from an unduly and improper application of the Algebra of Logic. The notion of finite definability is also rejected as a mathematical notion, just because it uses natural language. "Für den Cantorismus, aber gegen den Russellismus" is

[26] Exactly as the Liar, for which Peano refers to Russell's *Principles* [1903], but we haven't been able to locate the exact place; note that Russell's paper [1906a] is of May 1906; Peano's is dated August 23, 1906.

[27] That predicative definitions might be dangerous for mathematical practice is also held by Zermelo [1908a].

the final motto of [Schönflies, 1911], p. 255.

Schönflies anti-axiomatic and anti-formalistic inclinations naturally lead to consider Brouwer's early work. Brouwer's approach to paradoxes is to be found in the dissertation of 1907, chapter III ([Brouwer, 1907], cf. also van Dalen [1999], pp. 105). From a general perspective, we can summarize his position by saying that it is a straight consequence of his philosophy. Concerning Russell's contradiction, Brouwer (p. 162) notices that the usual logical principles only hold for words with mathematical content, not for linguistic systems, like those of Peano or Russell; for instance, in order to decide whether a class falls under a propositional function, the class ought to be completed. The contradictions show that there are propositional functions which define complementary (disjoint) classes and yet do not satisfy tertium non datur. Similar ideas can be found in the philosophical paper of 1908 *On the unreliability of Tertium non Datur* (see [van Dalen, 1999]).

There is an interesting idea that can be seen as a positive outcome of Richard's paradox in Brouwer's work.[28] Indeed, Brouwer [1907], p. 149 introduces the notion of *denumerably unfinished set*, i.e. a set in which we can determine only denumerable subsets of elements, but where these denumerable subsets do not exhaust the given set, so that we can immediately produce new elements of the set from any given denumerable such subset. Typical examples of denumerably unfinished sets are the totality of countable ordinals, the points of the continuum, and in particular, the set of all *definable points of the continuum* (ibid. p. 150), as it could be shown using Richard's paradox.

As to Burali-Forti, it is not a contradiction of mathematics, but it invests a logical structure (the collection Ω of all ordinals) with no proper mathematical content, since Ω is not considered by Brouwer a well-defined mathematical object, and, as such, it might be contradictory as well as consistent. Mathematically speaking, he suggests that the contradiction might be avoided by denying that the largest well-ordered type has a successor order type (p. 153; this is analogue to Bernstein [1905b]).

Among the French mathematicians, the semi-intuitionist Borel developed his own diagnosis concerning the Richard paradox. He has a paper on the paradoxes ([Borel, 1908]), where he 'solves' Richard's contradiction by introducing the distinction between *effectively enumerable sets* and *denumerable ones*. The set E of Richard is denumerable; this is because by finite means you can only determine at most a denumerable set of reals. Yet E is not effectively enumerable; i.e. you cannot define with finitely many words a procedure for assigning unambiguously a rank (i.e. a place) to each element of the set; you might eliminate the uncertainty only with "infinitely many words", but a definition requiring such means would be "en dehors des Mathématiques" (ibid. pp. 165). In order to make sense of the enumeration E as effective, *one ought to have solved all the mathematical problems which one could possibly state, because there are expressions which become definitions of a real only modulo a proof or the solution of a certain problem.*

[28] But Richard is not mentioned; by contrast, see *On the foundations of mathematics*, reply to a review by Mannoury; he credits Richard's paradox to Mannoury.

So Borel, like Poincaré adopts a point of view – definability in finitely many words – which is an extension of the algebraic Kroneckerian point of view: only objects constructible in finitely many steps are properly mathematical objects. However, unlike Poincaré, he does not even consider the problem of predicativity. According to Borel, all paradoxes of set theory derive from regarding as evident the following proposition:" Tout ensemble dénombrable est effectivement énumérable", which is false.

To conclude, we mention that at about the same time, there appears a rather long essay (in Italian) by Bebbo Levi *Antinomie logiche?* ([Levi, 1908]),[29] which deals with the antinomies of Richard, Burali-Forti, Russell and Cantor. While discussing the Richard paradox, Levi outlines an antinomy which is essentially a variant of Berry's paradox (though he does not mention Berry): let B be the number of symbols of our alphabet inclusive of exponentiation $\lceil$ and let B be written in decimal notation (say, with length β). Then the expression

"the number placed at position $B\lceil B$ in R"

has $36 + 2\beta$ symbols (R being a fixed ordering of the set of all expression of the language). Hence its place in the ordering is $< B^{36+2\beta} < B\lceil B$: contradiction.

Variations II: Hessenberg, Zermelo and the Grelling-Nelson paper

In this subsection we briefly report on further discussions which arise among set theorists on the subject of paradoxes; we also touch upon the philosophical paper of Grelling and Nelson.[30] This work (according to Bernays one of the best introductions to the subject) introduces a new paradox and gives an idea of the foundational debate in Göttingen; Grelling wrote his dissertation on a problem suggested by Zermelo and accepted by Hilbert, who also strongly supported Nelson and favoured his "kritische Mathematik", cf. Peckhaus [1990].

On the side of mathematical foundations, Gerhard Hessenberg's *Bericht* on the fundamentals of set theory ([Hessenberg, 1906]) devotes three chapters (ch.XXII: *Logical Completeness and decidability*; ch.XXIII: *The paradox of finite denotation*: ch.XXIV: *"Ultrafinite" paradoxes*) to a thorough analysis of foundational questions and contains interesting ideas about the philosophy of mathematics. For instance, he clearly stresses the distinction (p. 610) between those set theoretic definitions which *give criteria to decide effectively the membership to a given set*, and definitions which do not yield such methods (is there a general criterion to decide whether a given irrational number, like $2^{\sqrt{2}}$, is transcendental?)[31] He

[29] Of the same author, recall the 1902 paper *Intorno alla teoria degli aggregati* [Levi, 1902] where the axiom of choice is stated.

[30] Apart from the witness above, there is indirect evidence that, for some years, Husserl himself shared a certain interest in paradoxes; e.g. a dissertation by A. Koyré attempting a solution of the Liar, was rejected by Husserl [Zambelli, 1999]; Prof.Paola Zambelli (Florence) has kindly shown us some unpublished notes (Husserl Archiv, AI 35/4a-35/24a) of 1912, dealing with the foundations of set theory, where Husserl is apparently aware of the work done by Russell, Schönflies Zermelo, Weyl.

[31] Probably an implicit reference to Hilbert's seventh problem.

seriously consider Kronecker's constructive point against the arithmetized continuum;[32] but he holds that, although each irrational determines an infinite fraction and each infinite fraction possesses a formation rule ("Bildungsgesetz"), it is not true that such a rule is available through explicit finite means. Else, we should be forced to (a form of) *the paradox of finite denotation*, i.e. were formation laws to coincide with definable ones, the reals would be denumerable, against Cantor's theorem.

In dealing with set theoretic contradictions, Hessenberg speaks of 'ultrafinite Paradoxieen' (parr.96-99). 'Ultrafinite' is distinct from 'transfinite': this term must be properly predicated only of sets, to which the theory of transfinite numbers applies. This is not the case when you consider the Russell set or the set of all sets, or of all things, or of all ordinal numbers: they are ultrafinite as they go beyond the cardinality of every aleph. As to the proposed solutions, he notices that some contradictions, e.g. Russell's, are not dangerous for mathematics. On the other hand (p. 630), Burali-Forti's paradox and the set ON of all ordinals are mathematically well-defined. In the special case of the first antinomy, he rejects the solutions that ON is not a well-ordering (though each segment of ON is), or that ON is a well-ordering, but it doesn't have an order type. The point is rather that we can't add to ON an element m (so the successor is not defined on ON), i.e. *union is not defined* on the pair ON and m.

From a general philosophical perspective, Hessenberg tries to use a Kantian idea;[33] as in natural sciences the antinomies arise if Nature is conceived as a closed whole, similarly the collection ON and the set of all sets cannot be conceived as a completed totality. Thus the notion of 'ultrafinite' apparently suggests a theoretical approach, which is quite close to the limitation of size.

In spite of some criticism to the axiomatic method as a *via regia* to solve paradoxes,[34] Zermelo's axiomatization.[35] yields a sensible tool for sterilizing the contradictions, at least in set theory (the major foundational contributions being of course [Zermelo, 1908a], [Zermelo, 1908b]). So Burali-Forti's antinomy does not go through, since the collection of all order types does not exist as a set, and Russell's paradox simply becomes the theorem that every set M possesses at least one subset M_0, which is not an element of M and hence an argument to show that there cannot exist a universal set. Moreover, both the paradox of finite denotation (mentioned by Hessenberg) and Richard's are blocked in set theory, as the postulates ought to yield clear cut criteria for defining sets.[36]

When Zermelo elaborates the second well-ordering proof, he also attacks predicative definitions as a firm ground for mathematical existence against Poincaré, with a kind of indispensability argument: if one rejects impredicative definitions,

[32] Hessenberg sketches a formal analysis of Kronecker's idea that one ought to consider as meaningful only decidable disjunctions; at pp. 88-92; he explicitly introduces two 'modal operators' Δ, ∇, in order to distinguish decidable from undecidable sets.

[33] He is a member of the Fries school, together with Bernays, Nelson, Grelling and others.

[34] Recall Schönflies [1906] and 2.3.2 above.

[35] See Zermelo's paper of 1908, e.g. in [van Heijenoort, 1967]

[36] This is not entirely correct, as we know, till later developments

then one must reject standard mathematics, e.g. the Cauchy-Weierstrass proof of the fundamental theorem of algebra. Incidentally, the objection is not irrelevant: in the paper [Poincaré, 1909b] and in the 1910 lectures in Göttingen, Poincaré feels obliged to explicitly rebut Zermelo's criticism, and to argue that the fundamental theorem of algebra can be proven in a predicative way (ibid. p. 199).[37]

Grelling-Nelson's 1908 essay aims at a conceptual analysis of the different paradoxes, trying to set apparent solutions thereof apart. According to the authors, the paradoxes of Russell, Cantor, Burali-Forti have a common core, which can be enclosed in a general form (P):[38]

> "let M be the set of all sets and let M' be a subset of M; let $\phi : M' \to \Phi$ be a bijection. Then define $X = \{\phi(x)|\phi(x) \in x\}$, $Y = \{\phi(x)|\phi(x) \notin x\}$. If $Y \in M'$, then $\phi(Y) \in \Phi$. Assume $\phi(Y) \in Y$: then $\phi(Y) \in X$ by definition; but also $\phi(Y) \notin Y$: contradiction. If $\phi(Y) \notin Y$, then it would be an element of Y, by definition of Y: contradiction".

As a special case of (P), Grelling and Nelson produce a new paradox (credited to Grelling):

> "to each word there corresponds a concept, that the very word designates, and which applies to it or does not apply; in the first case, we call the word *autological*, else *heterological*. Now the word 'heterological' itself is autological or heterological. Assuming that the word is autological, the concept that it designates applies, hence 'heterological' is heterological. But if the word is heterological, the designated concept does not apply, so 'heterological' it is not heterological."[39]

It is worth mentioning that the authors distinguish between general presuppositions of the paradoxes (the principle of non-contradiction, Tertium non-datur and the "Dictum de omni et nullo") and special hypotheses, which involve ϕ (e.g ϕ is one-to-one, etc.). According to [Grelling and Nelson, 1908], restricting one of the general principles is not a viable route: one would reject the very basis for deductive reasoning (the analytical judgements). So the solution ought to conclude that such ϕ can't exist and to point out which special conditions (i.e. about ϕ) in the general form of the paradoxes are not met.

From a constructive point of view, the paper is not conclusive. It rejects different proposals (among them Frege, Russell's no-classes theory, Poincaré's predicativism), a main criticism being that if one tries to single out the guiding principles behind the solutions, one meets the very same difficulties one wishes to avoid. For instance, Poincaré's principle ([Poincaré, 1906b], p. 295) that all statements about infinite sets are not admissible is self-refuting: this very sentence is about infinite

[37] Intuitionistic proofs were soon given by Weyl and others. In present-day foundational terms, it is known that recursive analysis proves the theorem.

[38] The following formulation of (P) is a paraphrase from the German text. Let us recall that Bernays noticed that the reduction of Burali-Forti to (P) is defective.

[39] This is an almost literal translation from the German text.

sets, so is not acceptable. Similarly, if we restrict universal sentences to finite universes, the logical principles would not be universal and we encounter Liar-like difficulties in stating the underlying principles.

The remaining chance seems to state the principles with an exception clause: "all propositions of the given form ..., except this very proposition are not admissible". But in this case we have a proposition that refers to itself and we can proceed to: "All propositions of the given form with the exception of the proposition "all propositions of the given form with the exception of this proposition are not admissible" are not admissible". If you replace "this proposition" with its meaning, you obtain an infinite sequence of 'exception' propositions, so that to the claim of the theory we cannot attach any precise sense ([Grelling and Nelson, 1908], p. 115). In conclusion, it seems that Grelling and Nelson implicitly propose the idea that paradoxes are essentially related a sort of *infinite regress* (or divergence in basic logical operations, like substitution of equivalents).[40]

2.4 Russell's type theoretic approach, the late Poincaré and Weyl

Russell [1908] is a fundamental contribution not only for the theory of types, but also because it offers what might be called *the received view* about paradoxes. There we can find a comprehensive description of the contradictions, which includes: (1) the Liar (Epimenides); (2) the Russell paradox, also in its relational form (3); (4) Berry's paradox; (5) the König paradox of the least undefinable ordinal (see [König, 1905a], 2.2); (6) Richard's paradox; (7) Burali-Forti's antinomy. Most significantly, Russell develops a unified theoretical framework, which yields both a solution and a foundational doctrine, the theory of types, that is still alive nowadays and has descendants in logic and its applications. Since the Russellian theory of types is widely well-known and well-investigated in the literature, we only roughly summarize its basic features.

The basic observation is that universal quantification – understood as full generality, i.e. with x ranging 'over the whole universe'– does not make sense: when we state that $\forall x\phi(x)$ is true, we only claim the function $\phi(x)$ has the value 'true', for all arguments x for which it is meaningful. The essential point of the doctrine of types is that each propositional function comes equipped with a range of significance, i.e. a type, and that quantification is legitimate only over types. The paradoxes (or *reflexive fallacies*) teach us that certain collections, like the totality of all propositions, of all classes and so on, cannot be types. For instance, while we can quantify over the collection of men, we cannot properly say 'all propositions of the form $p \vee \neg p$ are true'; 'all' ought to be replaced by 'any' (i.e. by a *real variable*), and it should be understood as a schema, which produces a particular proposition, for each specific substitution of 'any proposition'. Moreover, when we consider "p is true" , what is meant, is not a definite propositional function of the propositional variable p, but different propositional functions according to the

[40] See the paradox of the exception, in [Bernays, 1973], Nelson's *Ges. Schriften*, bd.II pp. 207-8: All universal sentences except this very sentence are unreliable.

type of propositions that p has.

It follows that logical objects must be distributed in layers or types, which reflect different ontological complexity, (e.g. a propositional function must have a higher type than its arguments), and things satisfying a given condition can be collected into a type only if they form a collection of *homogeneous objects*. Formally speaking, this means that we cannot deal with general variables, but each variable must have a preassigned type.

This is not the whole story. As we know, the main idea to solve the paradoxes is the vicious circle principle: no totality can contain members defined in terms of itself. Concretely, the idea is implemented via a hierarchy of propositions (and more generally, propositional functions). So each type is ramified in *orders*, which ensure that an object is never defined by quantifying over a totality which counts the object itself as element. The lowest type consists of individuals (where individuals are entities without logical structure), while the second logical type embraces *the first order propositions*, i.e. those, whose quantifiers (if any) range over individuals. Then we can quantify over first order propositions, and we give rise to the third type, consisting exactly of second order propositions. In general, the $n+1$-type comprises propositions of order n, which contain quantification only up to $n-1$-order. Of course, since a propositional function can be obtained from a proposition φ by treating one or more of its constituents as variables,[41] the hierarchy of types and orders is naturally transferred to propositional functions, and it makes sense to speak of *order of a function* (its order being roughly the order of the value, i.e. a proposition, the function assumes when it is applied to an argument for which it makes sense).

Once the type theoretic structure is laid down, Russell proceeds to a solution of the paradoxes. The general diagnosis is that all contradictions presuppose the existence of a totality, which is not legitimate as a quantification domain in the theory of types. For instance, the Liar (1) refers to the totality of all propositions, since it can be rendered as " it is not true of all propositions p, that if I affirm p, p is true". But following the doctrine of types, we must replace in (1) 'all propositions' by 'all propositions of order n' and hence the Liar sentence is not asserting a proposition of order n, but of order $n+1$. Therefore the Liar is plainly false, without implying its truth and hence a contradiction; and this solves the paradox.[42] Similar arguments guide the analysis of (2)-(7).

The importance of Russell's paper of 1908 can be measured by its subsequent work in the foundations of mathematics, and by its technical developments which find their systematic development in the monumental *Principia Mathematica*, written in collaboration with A.N. Whitehead and published in 1910 (vol.1), 1912 (vol.2) and 1913 (vol.3). Russell [1906b] yields the first logical technical apparatus for handling predicativity. A crucial role is played by *predicative functions* of one

[41] Russell almost systematically confuses a propositional function with its defining formal expression.

[42] Compare with [Russell, 1906a], or 2.3.1.

argument,[43] i.e. those having as order the successor of the order of their argument; for instance, a predicative function of one individual variable must have order 1 (it is elementary definable in our present-day-terminology and it contains quantification only on individuals); similarly, a predicative function $\varphi(x)$ whose argument x has order 2 must have order 3 (hence it comprises quantification over order 2 at most). Now the famous axiom of reducibility AR states that *every propositional function is equivalent, for all its values, to a predicative function of the same variables* ; hence, in a sense, we can neglect the intensional aspects involved in the definition of a propositional function, and concentrate over the extensional ones. To this aim, predicative functions are enough: the rationale behind AR is that we have many ways of stating and proving a universal statement (one for each order), but all of them can be reduced to quantifications over predicative functions (and predicative functions can play the role of classes, i.e. canonical representatives of arbitrary complex concepts). Of course, AR is a basic tool for reconstructing classical mathematics.

The latest Poincaré's contributions of 1909-1910 ([1909b], [1910], [1909a]) are triggered by and can be partly seen as reviews of Russell [1908], Zermelo [1908b], [1909], and Schönflies [1909]. There are two main aspects that should be mentioned in connection to the paradoxes. First of all, in his contribution to *Acta Mathematica* (where he criticizes Schönflies supposed solution of Richard's paradox) and in his Göttingen fifth lecture, Poincaré provides us with an interesting restatement of Richard's paradox.

He observes that, after all, what is proven by Richard is that it is impossible to find a formula φ involving definable reals and natural numbers in such a way that every definable real corresponds to a natural number, according to the relation defined by φ: whatever formula we choose, we can always find a definable real which is not in correspondence to a natural number. Thus, for Poincaré Richard's paradox is a sort of refinement of Cantor's theorem, that we might have stated in the form: "there is no definable enumeration of definable reals".

It is also worth mentioning that in these papers the great French mathematician advances a new idea of predicativity, which, though informally sketched, is suggestive of later developments in definability and proof theory (see Feferman [1964] and [2005]). He no more insists on the vicious circularity of the definition involved in the contradictions; instead, he holds the view that a predicative classification is characterized by invariance, i.e. it cannot be disordered by the introduction of new elements; by contrast, impredicative notions are doomed to constant modification by the introduction of new elements. In a sense, Poincaré is hinting at some form of *invariance under extension* (that will be made precise by Kreisel [1960], via model theory and recursion theory).

At about the same time, Weyl's *"Habilitation"* lecture deals with *the notion of definition and definability in mathematics*, and it is directly relevant to the problems raised by Richard's paradox. Weyl [1910] addresses a typical metamathematical general question: when is a relation *explicitly definable* from a set of given

[43]The notion can be extended pointwise to functions of several variables.

primitive concepts in mathematics?

The problem is first tackled by investigating a concrete case, that of Euclidean geometry. On the ground of Pieri's axiomatization (1899), Weyl characterizes the explicitly definable concepts of elementary plane geometry as those that can be inductively generated by means of *five basic definition principles* from two primitive concepts, the identity between points $=$ and the ternary relation $E(a, b, c)$ = 'the distance of point a from point b is the same as the distance of point a from point c'. In present-day terminology, the definition principles correspond to closure under (i) the logical operations of negation, conjunction (Koordination), existential quantification (Fortnahme); (ii) the combinatorial operations of permutation and expansion (i.e. if we are given a relation $R(a, b, c)$, we also allow a relation $S(a, b, c)$ which holds if and only if $R(a, c, b)$ holds; if we are given a relation $R(a, b)$, we can also build a new relation $S(a, b, c)$ which holds if and only if $R(a, b)$). In other words, they reflect a *finite axiomatization* of the elementary comprehension principle, and they imply the existence of exactly those sets (relations), which are definable by formulas in the elementary language comprising two predicate symbols for $=$, E.

In the light of the geometric example, Weyl attacks the concept "durch endlich viele Wörter definiert" as not precise and not easy to grasp, but he observes that the very notion of 'definable' is built into the axioms of set theory, because of axiom III (Aussonderung), where Zermelo refers to the notion of 'definite Aussage'. Long before Fraenkel and Skolem, this notion is regarded as unsatisfactorily vague and Weyl's proposal is to make Zermelo's informal concept precise with the notion of 'relation explicitly definable from $=$ (extensional equality), $\in$ (membership) by means of the basic operations (or modifications thereof, more suited to the set theoretic context).

According to Weyl, the lesson we learn from Richard lies in the following distinction: on one hand, it is true that we are able to characterize only denumerably many subsets of a given set by means of explicit definition; but, on the other hand, new objects and (possibly uncountable) sets can be introduced by applying the remaining set theoretic operations, that are justified by the extant axioms, like power set or union. And this makes sense of Cantor's theorem and shows where the mistake is in Richard's argument.[44]

Weyl will come back to the same problem a few years later in *Das Kontinuum* [Weyl, 1918] with a different diagnosis. As in 1910, the set of sets of natural numbers that are available via admissible operations (to which now also a form of recursion is added) is denumerable; but Cantor's argument can be reinterpreted *as showing that in the given universe there cannot exist a relation yielding a parametrization of all subsets of natural numbers* (section 5)! So it seems that Weyl follows a relativistic attitude, according to which the extension of the universe of sets and their properties depend on the operations which are accepted

[44] As to the relation with the foundational issue "Is mathematics the science of $\in$ and of all relations that can be built up by means of the elementary principles above ?", Weyl is sceptical and holds that mathematics must keep its roots in intuition and physical application.

to construct sets.[45] It is to be mentioned that Weyl's attitude towards Grelling's antinomy is utterly negative: he considers it pure scholasticism (see [Weyl, 1910], section 1, p. 19): there is no way, according to him, of assigning a meaning to 'heterological'.

Rather ironically, a few years before, it was the Polish philosopher and logician S. Lesniewski, Tarski's "doctor- father", and a prominent member of the so-called Łwow-Warsaw school,[46] who presented in 1913 an eccentric formal analysis of the Liar, bearing some resemblance with scholastic solutions. Certainly this route can be considered peripheral with respect to the main stream of the period (namely the foundational debate), but we wish to recall it, since it is not unrelated with the fundamental contributions of Tarski, and with more recent contributions (Simmons [1988], Epstein [1992]).

In his paper *A Critique of the Logical Principle of the Excluded Middle* [1913], Lesniewski intends to show that the law of the excluded middle is not universally valid, but that it can be replaced by a suitable restricted form, which claims that if one of two mutually contradictory sentence (of the form 'a is b', 'a is not b') is false, then the other is true if and only if its subject is denotative (i.e. refers to an object) and its predicate is connotative (i.e. it means an attribute and refers to the class of all things falling under that attribute). In the course of his arguments (see [Betti, 1999]), he faces the problem of solving the Liar paradox, since it appears that this antinomy in the form of the sentence E

'The sentence stated by Epimenides at t_1 is false'.

falsifies its revised version of *Tertium non datur*, unless its subject is non-denotative (as the predicates 'false' and 'true' are considered trivially connotative). It is exactly at this point that Lesniewski hinges upon a semantical hypothesis, implying that the subject of E is non-denotative. The idea is that the paradox does not involve sentences (e.g. E), but specific occurrences of sentences, i.e *sentence tokens*: so, for instance, E should be properly written, say, in the form E_{e,t_1}, e being an index for "Epimenides". Once this is made clear, the point is whether the subject of E_{e,t_1} is denoting. The answer of Lesniewski is negative on the ground of the special assumption that *a sentence token cannot denotes itself.*[47] Thus it follows that the sentence which Epimenides utters at time t_1 is false, since its subject is non-denotative; but if I utter E at time t_2, E_{I,t_2} has a denotative subject (i.e. E_{e,t_1} and is true).[48]

[45] See also Poincaré [1909a]. This is implicit in a footnote of Weyl (1910), where he consider a solution of the continuum problem impossible without an exact formulation of the acceptable set construction principles; in the same place he also suggests a strengthening of Zermelo's system with a maximality principle, in the style of Hilbert's completeness axiom of geometry: no proper subuniverse will satisfy Zermelo's axiom ([Weyl, 1910], p. 304). Similar arguments are repeated in [Weyl, 1921].

[46] Today he is mostly known for his logical theories of parts and wholes (mereology) and for his radical nominalism; for more information on Lesniewski, cf. [Surma *et al.*, 1992].

[47] The assumption is more general and technical, but it amounts in our case to forbidding self-referential tokens.

[48] See Gaifman [1988], Simmons [1988]); in medieval logic, similar arguments are given by

To sum up, Lesniewski's solution can be classified as an indexical solution, where truth is applied to sentences-in-context (i.e. to specific occurrences or tokens) and where the type/token distinction is essential. On the other hand, the hypothesis rejecting self-reference for tokens is normative and makes clear that Lesniewski is struggling for a regimentation of natural languages. According to Lesniewski, paradoxes apparently teach us exactly this: natural language is at the same time too rich and too vague, for a development of a scientific semantics.

2.5 Circularity, sets and simple types

A sensible axiomatization of set theory, a systematic development of various forms of type theory, the introduction of important conceptual distinctions (like the dichotomy predicative/impredicative or the vicious circle principle): all these can be regarded as the most relevant by-products of the paradoxes in the first fifteen years of the century.

In the subsequent years up to 1930, the logic papers, which more or less touch upon the problem of paradoxes, contribute to give a definite shape to a strictly formal and mathematical notion of logical calculus (its final outcome being perhaps in the Hilbert–Ackermann textbook of 1928), to the clarification of the theory of types, with the definite appraisal of the notion of simple type versus the notion of order, and to significant enrichments of the notion of set.

An historical lesson to be learnt is how deep and wide is the influence of the Russellian approach for the ongoing and future foundational studies, and the concrete development of "symbolic logic";[49] at the same time, though less spectacular, the investigation of set theory as a proper mathematical theory attains a substantial progress not only on the axiomatic side, but also with the extension of set theoretic core notions towards unexpected realms.

To illustrate these points, we briefly review ideas of Mirimanoff,[50] von Neumann, Finsler, Hilbert, Chwistek, Ramsey, insofar as they are closely related to the issue of paradoxes.

Mirimanoff's papers ([1917a], [1917b], [1920]) are very suggestive: in these works the problem of self-membership for sets is neatly posed, and *circularity becomes, in a sense, an object of independent study in mathematical logic and set theory.*

The main aim of Mirimanoff is to address the fundamental problem of set theory: find necessary and sufficient conditions for the existence of a set. In this context, he introduces *a strengthening of extensional equality by means of a suitable isomorphism relation*, which is needed to deal with possibly circular sets and essentially corresponds to the isomorphism of the trees picturing the given sets.[51]

Ockham and others.

[49] Chwistek [1933] in his contribution to the Königsberg Conference of 1931 could write (p.368) that the simple theory of types is today a fashion and is used by the majority of logicians.

[50] A very clear and concise note, also including a short biography, about Mirimanoff's contribution is due to Specker [2001b].

[51] According to him [Mirimanoff, 1917a, pp. 40-41], two arbitrary sets E and E' are isomorphic if and only if 1. there exists a bijection $F : E \to E'$; 2. F is so defined that: 2.1 to each atom

He is then able to suitably rephrase Russell's argument in such a way that it naturally opens the route for non-well-founded sets. If we call a set *of first kind*, if it is not isomorphic with anyone of its elements, while a set is *of second kind* if and only if it is isomorphic with at least one of its elements, Russell's contradiction just shows that the collection R of sets of first kind does not exist as a set.[52]

A major achievement of [Mirimanoff, 1917a, p.42] is the fundamental distinction between *ordinary* (= well–founded) and *extraordinary* (= non–well–founded) sets: a set X is ordinary if every $\in$–descending chain in X is finite; it is extraordinary, otherwise (there exist infinite $\in$-chains). Clearly all second kind sets are extraordinary, but the viceversa is not true (for instance, consider the set $E = \{e_1, E_1\}$, where $E_1 = \{e_1, e_2, E_2\}$, $E_3 = \{e_1, e_2, e_3, E_4\}$, etc.). At p. 43, [Mirimanoff, 1917a], Mirimanoff states a generalization of Burali-Forti antinomy: the collection WF of ordinary sets (on a given set of atoms) is not itself a set.[53] He then introduces the notion of ordinal rank for ordinary sets and notices that ordinary sets can be arranged in a cumulative hierarchy, indexed by their ranks;[54] thus ordinary sets exist if and only if the collection of ordinal ranks of their elements is bounded.

Does the natural structure of ordinary sets weaken the ontological support for extraordinary sets? On the base of intuitive grounds, the answer appears negative. In [Mirimanoff, 1917b], p. 212-213, the author explicitly points to the use of extraordinary sets for modelling mirror-like situations: imagine a book whose cover contains a picture J where two children a, b are looking at the very same image J' of J on a smaller scale, where corresponding children a', b' look at the image J'' of J', and so on.... For instance, this could be represented by a set J satisfying the following equations:

$$\begin{aligned} J &= \{a, b, J'\}; \\ J' &= \{a', b', J''\}; \\ J'' &= \{a'', b'', J'''\}; \\ &\dots \end{aligned}$$

But then J is isomorphic to J' and hence $J = \{a, b, c, J\}$.[55]

Incidentally, it is worth mentioning that the distinction between ordinary and extraordinary sets is also used by [Mirimanoff, 1920], p. 33-34, in order to instantiate Poincarè's distinction between predicative and impredicative definition: J

of E F assigns an atom of E' and conversely; 2.2. F maps every set $e \in E$ onto a set $e' \in E'$, in such a way that (i) e' is the bijective image of e' (via F); (ii) each atom of e is sent onto a corresponding atom of e', and so on.

[52]Indeed, a set of second kind always contains a set of second kind; hence a set of sets of first kind must be of first kind. If R were a set, it should be of first kind; but then it could not contain all sets of first kind...

[53]This is also known as the paradox of grounded sets in the literature and it runs as follows: if WF is the set of grounded sets, then WF itself is grounded: were $WF \ni x_0 \ni x_1 \ni \dots$, then x_0 would be an ungrounded member of WF: absurd. Hence $WF \in WF$, so WF is ungrounded, contradiction. The same paradox is sometimes credited to Shen-Yuting [1953].

[54]Is this the first published and explicit version of the cumulative hierarchy?

[55]This statements is understood on merely intuitive, motivational grounds.

could be classified as impredicative, as it involves in its very definition a collection (J itself), which it is a member of.

Mirimanoff[1917a] also presents a sort of axiomatization, comprising, together with union, power, subset and urelements, (i) an axiom of "Burali-Forti", yielding a sufficient condition for sethood in terms of size, i.e. if a set is in bijection with the set of all ordinals, then it does not exist,[56]; (ii) a replacement axiom claiming that, if a set E exists together with a bijection φ which associates to each member of E a well-founded set, then the then $\{\varphi(a)|a \in E\}$ is a set. Mirimanoff's postulates do not include an explicit form of the axiom of foundation, nor any form of anti-foundation in our sense (Aczel [1988]). As far as the above axiom of Burali-Forti is concerned, it would be enough to require an application onto the ordinals, as a condition granting non-sethood; however, this axioms prepares the route to von Neumann's celebrated characterization of classes and to a beautiful application of the Burali-Forti antinomy.

Without touching upon von Neumann's system of 1925, let us briefly recall that his system assumes two sorts of objects: objects of type II (functions, corresponding to classes) and objects of type I (arguments), which are linked by the application operation of a function to its arguments. The two domains partly overlap and there are objects of type I-II, which correspond to sets (being at the same time functions which can also be arguments). The fundamental axiom IV-2 of von Neumann's system states that an object a is a proper class (i.e. it is not type I-II) if and only if the totality of its members can be mapped onto the totality of all arguments i.e. onto the universe of sets.[57] Then the Burali-Forti antinomy becomes a lemma (the class ON of all ordinals is not a set), which implies with axiom IV-2 that there is an application of ON onto the universe of all sets, and hence that *the universe of sets is well-ordered*. Thus von Neumann characterizes classes as distinct from sets by means of a mathematical property, and this settles the problem of making 's distinction between inconsistent and consistent precise and applicable. At the same time, it shows that global choice can be made a theorem under a suitable view of sets.

Let's go back to circular sets, which are dismissed by von Neumann as superfluous (he is aiming at a hierarchical model of set theory). Like Mirimanoff, Finsler's paper [1925] *Gibt es Widersprüche in der Mathematik?* properly belongs to the foundations of set theory (so we won't go into details),[58] but it deserves to be mentioned, since non-well founded sets are again directly and explicitly motivated by the attempt of understanding the roots of the antinomies. According to Finsler, the point is that paradoxes hinge upon circular notions. However, not every circular definition leads to a contradiction; there are circular definitions that are not satisfiable, but some of them are admissible, in the sense that they ensure the existence of the defined object.[59] In particular, Cantor's general notion of set

[56]In Mirimanoff[1917a, pp. 45–46],[1917b, pp. 213-15] we can find *in nuce* the idea of von Neumann ordinals cf. von Neumann [1923], [1925].

[57]All this can be properly rephrased in the language of functions and arguments.

[58]For a short biography and summary of Finsler's work, see Specker [2001a].

[59]He mentions that $x = a - x$ and $x = a + x$ are both circular, but the first, if a is a fixed real

is regarded as intrinsically "zirkelhaft": one allows sets that depend on sets or on general things depending also on sets (p. 151), and nothing ensures that this dependence chain again drives us back in a circle.

From a technical point of view, an original intuition of [Finsler, 1926b] is *the use graph theory for representing circular structures*. So he suggests to use arrows for interpreting membership, and it is easy to figure out a set that has itself as unique element (p. 692) and more complicated circular situations. Interestingly enough, a circular set H containing itself as element will correspond to an equation of the form $H = \{A, \ldots, H\}$ (p. 693).

Like Mirimanoff, he also realizes that, once reflexive sets are present, e.g. auto-singletons satisfying $a = \{a\}$, $b = \{b\}$, equiextensionality is not enough as identity condition and one has to introduce an appropriate notion of isomorphism (that is, a form of bisimulation, in our sense, see [Aczel, 1988]).[60]

We wish to conclude on Finsler by mentioning his 1926 paper *Formale Beweise und die Entscheidbarkeit*: there he applies the machinery of Richard's paradox, in order to produce metamathematical results, in particular 'formally undecidable propositions'. Finsler's arguments do not conclude to a proper anticipation of the Gödel incompleteness theorems[61]; but they show that an attentive reading of the paradox might have unexpected applications.[62] Finsler's approach is informal and rather vague; in particular, he holds the view that it makes sense to figure out an absolute consistency proof against a formal one; he further opposes formal results to conceptual results, where 'conceptual' is left unexplained and it corresponds to his own understanding of 'metamathematical', which is apparently far away from Hilbert's sense.

Coming to the development of logic, Hilbert's lecture notes for the Winter–Semester course 1917-1918 *Prinzipien der Mathematik* [Hilbert, 1918b] open an inward look into the foundational ideas of the great mathematician. Besides a formulation of first and second order logic (which in essence anticipates Hilbert-Ackermann's exposition of 1928), the text outlines a formal extended functional calculus (p. 188) where functions and propositions occur also as arguments, and unrestricted comprehension (in the form of substitution rules) is accepted, among other formal rules. It is then shown how to develop set theory in this extended logic, and how to formally derive in full details Russell's predication paradox, and

$\neq 0$, uniquely fixes a real number; the second has no solution.

[60] Finsler's ideas are displayed into an axiomatic proposal, mainly based on a completeness axiom *á la* Hilbert.

[61] On the limit of his ideas, see the discussion in [van Heijenoort, 1967], 438-440.

[62] For further discussion on paradoxes, see Finsler and H.Lipps 1927, Phil.Anzeiger II, Jg.1927.

a suitable formulation of the Liar [63] and a variant of the Berry's antinomy.[64] Hilbert's diagnosis is deeply influenced by Russell: the problematic assumption is located in the admissibility of predicates and propositions as objects, i.e. formally in the expressions $X(Y)$, $P(P)$, and finally in circularity and impredicativity. Thereafter, Hilbert describes the ramified theory of types with the reducibility axioms, showing how certain non-trivial parts of mathematics can be developed by means of it. Note that *is not Zermelo's theory that is lectured upon, but Russell's.* Historically, it is interesting that in the discussion of paradoxes given ten years later in [Hilbert and Ackermann, 1928], Hilbert, in cooperation with Ackermann, sticks to the *Principia Mathematica*, in order to formalize the paradoxes; the discussion of the Liar given in Hilbert-Ackermann [1928] (pp. 157-163) is essentially the same described in the 1917/18 lecture notes.

Of course, Hilbert is appreciative of Zermelo's theory as well: in his paper of 1918 *Axiomatisches Denken* ([1918a]) Zermelo's axiomatization is mentioned as a contribution, vindicating set theory from contradiction. His positive attitude towards Zermelo is even more transparent in the Bologna lecture of 1928 *Probleme der Grundlegung der Mathematik* , where he criticizes the ideology behind Russell's theory (as supporting the prohibition of impredicative propositions), and regrets that Zermelo's path is abandoned under the pressure of 'influential circles' (the constructivistic school). However, it is significant that proof theory begins as a work on (a fragment of) Russell's theory of types, a fundamental metamathematical theorem concerns *Principia Mathematica* (Gödel [1931]), and Zermelo's theory is not directly considered in the 1928 textbook in Mathematical Logic.[65]

That the discussion on paradoxes has still a role to play in the proper evaluation of type theory and in disentangling its specific technical potentialities, can be ascertained from the work of Chwistek and Ramsey.

Chwistek [1921] deals with the antinomies of formal logic, and he intends to criticize *Principia Mathematica* in two directions. First of all, he holds the view that Russell's theory is not sufficient to avoid Richard's classical antinomy. The specific analysis of paradoxes, given by the author, appears unconvincing, though dressed in the type theoretic formalism. For instance, he claims that, since there exists an enumeration $\{\varphi_n(x)\}_{n\in\omega}$ of all definable[66] propositional functions, where x has, say, the type of the natural numbers, and the functions can be assumed

[63] It is a new variant of the traditional Liar: Hilbert introduces a predicate $Beh(X)$ for " X is asserted by K in the time interval t", A stands for "whatever K asserts in t, is false". Then, assuming

1. $Beh(A)$;
2. $\forall X(Beh(X) \rightarrow A \equiv X)$,

a contradiction is derived as usual.

[64] In a finite amount of time, e.g. in the 20th century, only finitely many numbers can be defined in the formal language Hilbert has outlined; then consider "the first number not defined in the 20th century"; this leads to a contradiction.

[65] The influence of Russell's ideas on Hilbert's school is further analyzed by [Mancosu, 2003].

[66] In the type theoretic language.

predicative by the axiom of reducibility,[67] then we can construct (= define) a naming relation R such that $R(n, y)$ holds if and only if y is the class of all objects satisfying $\varphi_n(x)$. But if this were the case, of course, by diagonalization we might easily obtain a contradiction.

As to the Liar paradox, Chwistek shows that a version thereof can be reconstructed in the simple theory of types without the axiom of reducibility, once we assume the existence of a proposition A and of a predicate M of propositions such that[68]

$$\forall p(M(p) \leftrightarrow (p \leftrightarrow A))$$
$$A \leftrightarrow \exists q(M(q) \wedge \neg q)).$$

The solution proposed by Chwistek is prompted by a constructive/predicativistic conception: A involves quantification over all propositions, it is not a proposition stating a directly given fact and hence there is no reason to assume its existence. Or else, we might consider the above formula as an axiom, but then we have to show its consistency and the contradiction shows exactly this is not the case.

What is more interesting, is that, according to Chwistek, Russell's theory is not necessary, at least if one sticks to the consideration of set theoretic (logical) paradoxes. The simple theory of types is sufficient: for him, as a few years later for Ramsey, the introduction of predicative functions and of the axiom of reducibility are simply complications. Similar arguments are contained in [Chwistek, 1922] (*Über die Antinomien der Prinzipien der Mathematik*): if Richard's antinomy belongs to logic and not to semantics, Russell's type theory is not enough to save us from contradiction; if it belongs to semantics, then the theory is superfluous: solving Russell's paradox calls for the simple theory of types. On his side, Chwistek adheres to a kind of nominalistic position, and he tries to develop a theory of constructive types where it should be possible to develop to a certain extent the foundations of analysis (e.g. to define Lebesgue measure) without assuming the existence of the set of real numbers (in coherence with Poincarè's inspiration).[69]

Ramsey ([1926]) develops *the celebrated distinction between logical and epistemological contradictions* (Group A vs. Group B). Logical contradictions only involve mathematical or logical terms, like class, number, and hence show that our logic or mathematics is wrong. Semantical contradictions cannot be stated in purely logical terms and involve notions like 'thought', 'language', 'symbolism', which, according to Ramsey, are empirical (not formal) terms. Hence they are due to faulty ideas concerning thought or language; they pertain to "epistemology".

[67] See the criticism of Ramsey [1926], footnote 27, to this step. According to Ramsey, the definability of a function by means of a fixed set of notions, does not imply that there exists an equivalent predicative function.

[68] Informally, $M(p)$ is to be read as meaning "I consider the elementary proposition p as interesting", so that the first statement claims that "I consider as interesting exactly those propositions equivalent to the proposition: *I consider as interesting a proposition and this proposition is false*".

[69] Chwistek's attempt of founding mathematics without the axiom of reducibility is termed "heroic" in the introduction to the second edition of the Principia Mathematica, 1925.

For the solution of the antinomies of the first group, Ramsey relies upon a universe of mathematical objects which is structured in types: individuals, functions of individuals, functions of functions of individuals, etc. Quantification on arbitrary types makes sense and hence types are closed under impredicative comprehension, which is necessary for mathematics and quite legitimate (see, by contrast, the position of Chwistek). Types reflect an intrinsic characteristic of the logical and mathematical objects (the ontological complexity, see p. 372) and the logical paradoxes are exactly those which require type distinctions to be solved (e.g. self-membership is blocked for objects in the type theoretic hierarchy).

Of course, Ramsey is left with the problem of solving the semantical antinomies. In accord with his mathematical realism and against to Russell's ramified approach, he rejects the axiom of reducibility and accepts a liberalized notion of predicative function (pp. 367–368). The class of predicative functions is, in his view, closed under quantification, in the sense that a predicative function of individuals allows for quantification on individuals, a function of functions of individuals may also contain quantification on functions of individuals, and so on. But then he has the problem of avoiding e. g. Grelling's paradox (p. 358) without using the order distinction. Indeed, if we define the notion of "heterological", $Het(x)$, as $\exists\varphi(xR\varphi \wedge \neg\varphi(x))$, where we apply a given notion R of meaning and the quantifier range over predicative functions (in extended Ramsey's sense), we have that Het is again a predicative function. Hence, if we proceed to apply R to Het, we do regain the paradox.

Ramsey's solution is rather obscure and sketchy. First of all, he proposes to distinguish several notions of meaning. In the light of later developments, it is interesting to read that for him *semantics has no viable universal notion*: in particular, "it is impossible to obtain an all-inclusive relation of meaning for propositional functions. Whatever one we take there is still a way of constructing a symbol to mean in a way not included in our relation. The meanings of meaning form an illegitimate totality" ([Ramsey, 1926], p. 372). So, in the specific case of defining the notion of "heterological", he argues that the starting meaning relation R cannot be used, and one needs a new meaning relation depending on the given fixed R.

Even if these ideas are reminiscent of the soon–to–come Tarskian ideas, Ramsey is very far from the neat distinctions which are preliminary to Tarskian semantical investigations and he is still entangled in a confusion between functions and their symbols. In the end, he resolves the semantical antinomies by restoring the order distinction (p. 373), even if in his logic the hierarchy of orders is no more in accord with Russellian predicativistic views.[70]

[70] For an analysis of semantical antinomies in ramified context, compare also the later contribution of Church [1976].

3 BETWEEN METAMATHEMATICS AND TYPE FREE FOUNDATIONS: FROM 1930 TO 1945

In this section we are concerned with the influence of paradoxes in two distinct directions, involving (i) proper metamathematical investigations, but also (ii) ideological foundational aims.

3.1 A machine for paradoxical constructions

Let us first consider (i): its main by–product is a flexible arithmetically grounded technique for taming self-reference and the foundations of scientific semantics. In the metamathematical investigations of the thirties paradoxical arguments transfigure into specific results, which grow out of the strict formalization of the syntax and the semantics: firm grounds are reached for simulating (indirect) self-reference.

Indeed, while Gödel intuitively and explicitly relates his construction of formally undecidable sentences to epistemological paradoxes ("the analogy with the Richard antinomy leaps to the eye", [Gödel, 1931], [van Heijenoort, 1967], p. 599), the core of his work lies elsewhere,[71] i.e. in the technical exploitation of number theoretic techniques (e.g. the "Chinese remainder theorem") for the analysis of syntactical substitution.[72] The legitimacy of self-reference is reduced to the fact that there exist natural languages and theories (e.g. fragments of *Principia Mathematica*, number theory, etc.), which can construct a mathematical interpretation of their own syntactical apparatus, and, in particular, of the concepts of provability, (partial) truth, etc. One can figure out plenty of self-referential statements, but *self-reference is to be understood in the indirect sense*: we may well have formulas $\varphi(x)$, which express properties of their own "name" $\ulcorner\varphi\urcorner$; but no vicious circularity arises.

It was soon realized that Gödel's construction could be given a general form, i.e. *paradoxes and limitations of formal systems follow from a general simple lemma, which did not quite refer to provability or truth, but to arbitrary definable properties.* This is to be found already in Carnap's *Logische Syntax der Sprache*, p. 91 ([1934b]; also [1934a], p. 270), and a few years later in Rosser's paper of 1939:[73]

> for every formula $\psi(v)$ with only v free, there exists a sentence φ such that $\varphi \leftrightarrow \psi(\ulcorner\varphi\urcorner)$ is provable.

The lemma yields a kind of machine for producing self-referential statements as it can easily be applied to arbitrary concepts, which are expressible in a sufficiently rich language. For instance, Carnap [1934a] consciously applies the diagonalization property, as a uniform tool for transforming the semantical paradoxes into

[71] Contrast this with Finsler's argument [1926a].

[72] That substitution could be applied in order to produce a contradiction, was already known: see [Pelham and Urquhart, 1994] and the Russellian paradox killing the substitutional theory.

[73] See [Rosser, 1939], p. 57, Lemma 1. For a survey, see Smorinski [1981]

undefinability and (formal) undecidability results.[74]

The range of its applications and the recondite algebra underlying the Gödelian construction will be grasped only much later during the 1970's. A few years later a suitable analog of the diagonalization lemma (the so-called second recursion theorem) was discovered by Kleene, and was soon to become a basic tool in the foundations of recursion theory and computability theory.

3.2 Tarski's approach to semantical paradoxes

As we have seen, the direct impact of *logical paradoxes* for the foundations of mathematics was largely sterilized towards the end of the first decade of the century: mathematical arguments were discovered, that solved or explained the contradictions away.

By contrast, it is evident from most significant papers of the twenties, that the question of finding a formal solution to the *semantical* contradictions remained open. At best, one could propose Russell's theory of ramified types as a means of solving the Liar; but the route had become much involved, due the Richard paradox, and (even in Ramsey, Hilbert, Chwistek) it had not been pursued beyond the level of an informal hint or a short formal sketch. No detailed formal analysis of semantical notions (like truth or definability) was ever explicitly undertaken, which could be legitimately regarded as a proper *mathematical solution* to the semantical side of paradoxes. Of course, the reader might ask why this would be worth studying from a logical-mathematical point of view, why semantical investigations could possibly have applications other than purely methodological or philosophical.

The answer is not clear-cut or simple, but it should not be underestimated that semantical notions — in particular the notion of definability — were more or less explicitly used in certain parts of set theory (descriptive set theory), or "set theoretically inclined" parts of function theory (e.g. classification of measurable sets and functions). After all, it should not be forgotten that Richard's antinomy and its variants emerged along with the debate on the arithmetized continuum and the related well-ordering problem (recall 2.2).

In addition, certain foundational views (especially those arising from French semi-intuitionism or from Poincaré's reflections) apparently required a precise notion of definability; and definability (in a fixed language) might possibly be regarded as a kind of generalized constructivity.

Polish mathematics was in the twenties highly advanced in set theory, topology and function theory;[75] at the same time, the idea of a formal methodology and in particular of a scientific semantics, was intensely developed by prominent Polish philosophers and logicians working in Łwow and Warsaw ([Wolenski, 1995]), among them Lesniewski, Łukasiewicz, Chwistek, who did not disdain cooperation with

[74] Was Carnap aware of Tarski's undefinability theorem of [Tarski, 1936]?

[75] Just remind the names of Sierpinski, Kuratowski, Banach, etc.

mathematical colleagues; after all, the journal *Fundamenta Mathematicae* had Lesniewski among its founders.

It is this unusual and stimulating environment, that forms the background to Tarski's analysis of semantical paradoxes.[76] As evidence for this, one should not forget that Tarski worked on the mathematical side of the problem : the semantical investigation is preceded by the note ([Tarski, 1931], *Sur les ensembles definissables des nombres réels I*, and, jointly with Kuratowski, *Les opérations logiques et les ensemples projectifs*, [Kuratowski and Tarski, 1931]), where logical notions are explicitly used to define certain collections of mathematical objects (e.g. to characterize sets of sequences of reals which can be represented as finite sums of elementary linear sets) and to establish closure properties of Lusin's projective sets. In the introduction to the first note, he explicitly mentions that he wishes to avoid the distrust of mathematicians towards definability, obviously ascribed to the arising of antinomies (Richard's is mentioned). Consequently, he sticks to a strict mathematical style, fixing the logical system of simple type theory with extensionality as the main tool for characterizing definability in unambiguous manner (Chwistek, Carnap are his references).

Coming closer to Tarski's monograph, its first basic achievement is that it clearly specifies a precise formal requirement to be met in the semantical investigation of truth. Its main goal is to provide "a materially adequate (*sachlich zutreffende*) and formally correct definition of the term "true sentence" (*wahre Aussage*)". What this properly means, is carefully discussed in the first section and explained by means of the celebrated schema (T):

x is a true sentence if and only if p

where p stands for a sentence and x is a name of p. The naive idea is that, if we have a definition of truth according to the classical correspondence intuition,[77] then we must be able to justify that *every* instance of the schema (T) holds.

Then Tarski observes that, though quite natural, we cannot rely upon (T) as a definition of truth, since the Liar paradox (Tarski adopts a version credited to

[76] We won't touch the detailed content of Tarski's monograph [1936], which properly pertains to the history of model theory; nor we touch the influence of Tarskian ideas upon philosophy. Concerning his direct debts, Tarski owes inspiration to his teacher Lesniewski for semantical ideas. Also Chwistek attempted an *elementare Semantik* in the years 1926–1933 ([1926], [1929], [1932]) on the ground of a nominalistic foundational program. He aimed at a reduction of Logic and Mathematics to a kind of "rational semantics", which he believed to be quite close to Poincarè's ideas [Chwistek, 1932]. Chwistek's semantical approach is extremely formalistic and far from the Tarskian ideas; according to his view, the fundamental notion of semantics is the substitution relation "H is the result of the substitution of G for F in E" ([1933], p. 374).

As to the dates, Tarski specifies that his work was presented by Łukasiewicz to the Polish Society of Sciences in Warsaw already in March 1931, but the results originated in 1929 and were reported in two communications of October and December 1930.

[77] A true statement is a statement making a certain claim about the world, and the claim actually holds; this conception is to be found already in Plato, *Soph.263a*, Aristoteles, *Met.* Γ*, 1011b, 25-30.*

Łukasiewicz) immediately arises.[78] Of course, it's easy to ascribe the source of the contradiction to the fact that the sentence substituted for p in the schema (T), in order to get the antinomy, contains the term "true". But, as Tarski remarks, it is not clear why we should restrict (T) to sentences not containing "true".

From a technical point of view, leaving the problem of contradictions aside for a moment, one is naturally driven to characterize the true sentences by a single axiom which implies all instances of the schema (T); but this, as Tarski observes, leads to accept quantification over sentences[79] and a quotation functor (Anführungsfunktion) $p \mapsto \lceil p \rceil$, which sends sentences into sentence names. Then (T) could be rephrased as

$$\forall x(x \text{ is true} \leftrightarrow \exists p(x = \lceil p \rceil \wedge p)).$$

Already the quotation functor is dangerous: Tarski observes that we can reconstruct the Liar paradox without even mentioning the term "true'. Simply stipulate that

$$c = \lceil \forall p(c = \lceil p \rceil \rightarrow \neg p) \rceil$$

and accept the obvious law

$$\forall p \forall q(\lceil p \rceil = \lceil q \rceil \rightarrow (p \leftrightarrow q));$$

standard logic readily leads to a contradiction.

Given these obstacles, Tarski proceeds to consider a *structural definition*, i.e. a definition which essentially exploits the logical form of an expression and the fact that expressions are recursively defined: for then we can specify that a true sentence is (i) a sentence which possesses certain structural properties (i.e. properties concerning the form and the disposition of the single parts of the expression), or (ii) a sentence obtained from structurally described expressions by means of structural rules (e.g. if an expression γ has the form $\alpha \wedge \beta$ and α, β are true, so is γ).

But this route is only viable for a language which is structurally described and completely described, i.e. a formalized language. So it is not the case of the natural language with its universalism. Paradoxes definitely show, according to Tarski, that semantical investigations must concentrate on formal languages.

The informal conclusion that Tarski draws from the Liar is that there is no interpreted language which is free from contradictions, obeys to the classical laws of logic and satisfies the requirements (I)-(III), where

1. (I): the language has names available for all its sentences:

[78] Let c be a typographical abbreviation for the sentence: " c is not a true sentence". So by stipulation, c must be identical with "c is not a true sentence", and the schema (T) yields:

" c is not a true sentence" is true if and only if c is not a true sentence, whence by substitution of identicals, we get the contradiction: c is a true sentence if and only if c is not a true sentence.

[79] Remind that Lesniewski had investigated a system of second order propositional logic, the so-called protothetic.

2. (II): any expression obtained from (T) by substituting p with an arbitrary sentence of the language and x with a corresponding name of p is accepted as true in the language;

3. (III) there exist self-referential sentences, i.e. it is legitimate that a sentence comprises its own name as constituent (so that we can legitimately stipulate that the name c denotes a sentence of the form $\alpha(\dots c\dots)$.[80])

This is only a rough summary of the first section of Tarski [1936]. The subsequent sections lay down the mathematical foundations of formal semantics and we only touch briefly upon a few points, which are explicitly related to the question of paradoxes. Conceptually, the Tarskian achievements can be split into a sequence of main steps.

A. Tarski introduces the distinction between the formal language itself whose semantics we are after, and the language used to talk about it, the so called *metalanguage*. In the metalanguage we typically have names of all expressions (terms, formulas,...) of the object language, but we also talk about the meaning of the expressions of the language L. In particular, to every sentence φ of the object language L there are associated (i) a sentence φ^L translating φ in the metalanguage; (ii) a name $\lceil\varphi\rceil$ in the metalanguage of φ. This leads to a more accurate formulation of the (T)-schema: a formally correct definition of the symbol "Wr" (true) for the language L in a given metalanguage for L is an adequate definition of truth if and only if it implies
1. every sentence we obtain from" $x \in Wr$ if and only if p", by replacing x with a name (in the metalanguage) of a sentence φ of L and p with φ^L, i.e. the meta-linguistic translation of φ
2. the sentence "for all x, if $x \in Wr$, x is the name of a sentence".

B. In parallel with a mathematical development of the theory of formal languages, Tarski establishes the foundations of a "methodology of deductive sciences", introducing deductive systems, their provability relation, and their basic properties, like completeness, consistency, etc. Formal languages, deductive systems and their basic notions can be interpreted in number theory (via the technique of arithmetization, p. 301, sec.2 [81]) and their study necessarily leads to consider infinite collections. *Therefore the use of inductive definitions and consequently of inductive proofs has a prominent, essential role in metamathematics.*

C. Since there are infinitely many sentences, one is forced to a recursive definition of truth; however, since induction would break down when defining the truth of quantified sentences, the central notion of semantics eventually becomes the satisfaction relation "the sequence f of objects satisfies the formula φ " By means of satisfaction, one can easily define the notions of definability, denotation, truth, logical consequence (relative to a given object language).

[80] Tarski observes that (III) is superfluous if one looks at the Grelling antinomy.
[81] Tarski does not refer here explicitly to Gödel.

Concretely, the project is implemented by Tarski with the definition of satisfaction for a special first order language (the language of the calculus of classes, where one has the predicate of inclusion, the constant 1 for the universal class, and the operations of complement and union), and it is then shown (this is a metatheorem) that (i) the definition of truth via satisfaction is materially adequate; (ii) it has the expected properties of classical truth (principles of non-contradiction, tertium non datur); (iii) the true sentences of the language of class theory are exactly the theorems of class theory (Completeness Theorem for the calculus of classes).

D. Once we have the new semantical machinery, the problem arises whether we can find a formal counterpart of a universal language, where it would be possible to define an adequate notion of truth for the same language. The theory of simple types,[82] which comprises the axioms of infinity and extensionality, looks a promising candidate as general metatheory: it works as a general morphology for formalized languages, and it deals with arbitrary complex semantical categories (as we have means to talk of arbitrary higher types). However one has the fundamental negative result (see p. 370):

> " whichever definition we choose in the theory of types for the term "Wr"(= true), then it is possible to deduce in the same theory the negation of some instance of the adequacy schema (T).

In the proof of this theorem, Tarski (cf.footnote 88) exploits the technique of arithmetization and diagonalization, closely following the Gödelian pattern (assuming that we have in the metatheory an arithmetical definition Wr of truth, we can construct an arithmetical sentence ψ such that $\neg Wr(\lceil\psi\rceil) \leftrightarrow \psi$; if Wr satisfies the (T)-schema, we have a contradiction).

On the positive side (theorem II, p. 379), if we fix k, it is possible to construct in the theory of types an adequate definition of truth *for all sentences of the theory of types which contain variables of type at most* k.

More generally, the concept of truth can be adequately defined for any formalized language L, under the condition that our metalanguage is of higher order than L (so if L is the theory of simple (finite) types, we should need at least a theory with a transfinite type); and that the hypothesis is necessary, follows from the fact that we cannot define an adequate concept of truth for L in a metalanguage having order at most the order of L itself.

In the final section Tarski envisages the possibility of developing an axiomatic theory of truth (p. 380, theorem III); indeed, the semantical constructions show that, if the theory of types is consistent, it is also consistent to extend type theory with a primitive truth predicate V satisfying the adequacy conditions for the type theoretic language (without V). The resulting theory is not entirely satisfactory (e.g. how to derive the principle of non-contradiction for V in its full generality?); so the theory could be supplied with axioms expressing – besides non-contradiction – the closure conditions of truth. But Tarski criticizes this approach (one ought

[82]The type 0 is the type of individuals; the type n+1 is the collection of all classes of type n objects.

to have an argument ensuring the axioms uniquely characterize truth, but this would require some strengthening), and he even considers an extension of the formalism with the so-called ω-rule. We can't go into further details, but it is worth remembering that just the investigation of axiomatic theories of truth is far from trivial and rich of results (see §6).

The importance of Tarski's work is difficult to overestimate. He immediately realizes that (p. 401) his own semantical methods prompt him a way to show that Gödel sentences are true and yet formally undecidable. Also, he can make sense of Gödel (1931, footnote 48): "the true reason of the incompleteness is that the formation of ever higher types can be continued into the transfinite[...], while in any formal system at most denumerably many of them are available". Tarskian semantical notions play exactly the role of the higher types hinted at by Gödel. The argument (presented in the postscript) is quite simple. Choose γ such that $\gamma \leftrightarrow \neg Bw(\lceil\gamma\rceil)$; then in a metatheory which is strong enough to develop semantics, we have:

$$Wr(\lceil\gamma\rceil) \leftrightarrow \gamma$$

Hence $Wr(\lceil\gamma\rceil) \leftrightarrow \neg Bw(\gamma)$. If the theory is correct, we also prove in the metatheory

$$Bw(\lceil\gamma\rceil) \rightarrow Wr(\lceil\gamma\rceil)$$

which immediately implies $\neg Bw(\lceil\gamma\rceil)$ and $Wr(\lceil\gamma\rceil$, i.e. the Gödel sentence is unprovable, but true. Also, since $Wr(\lceil\gamma\rceil) \rightarrow \neg Wr(\lceil\neg\gamma\rceil$, and $Bw(\lceil\neg\gamma\rceil) \rightarrow Wr(\lceil\neg\gamma\rceil)$, we have $\neg Bw(\lceil\neg\gamma\rceil)$. So, even if Tarski tends to emphasize the distance between his own investigation and Gödel's, his conclusive observation achieves a unified view, showing the interplay of the semantical notions with the syntactical ones. The incompleteness phenomenon is better understood with the help of semantics. To sum up, Tarski's work eliminates the semantical notions, in favor of the (extensional) notions of type or set, and a theoretical explanation of semantical paradoxes, like the ancient Liar or Richard's, is finally achieved as a by–product of a systematic theory.

3.3 Type free foundations ?

One might say that in the twenties and in the early thirties the orthodox view among mathematical logicians was becoming more or less type-theoretic or set theoretic. Nevertheless, unorthodox theories were devised and investigated at length, that could be characterized as *non-extensional* and *operation-based*. In particular, there is an effort to elaborate grand logics, as a reaction to the logic of *Principia Mathematica* (both on philosophical and technical grounds).

This is partly motivated by the attempt to recover the original simplicity and the flexibility of the type free approach, as derived from the so-called naive comprehension principle (both on the extensional and intensional interpretation), but also by metamathematical needs, e.g. the demand for a clarification of fundamental concepts underlying the notions of 'formal system', 'formalism', 'rule', etc.

The idea that a type–free approach could be viable, is still alive and experiments are done assuming as primitive the notions of (self-applicable) function-in-intension (operation) and property (or predicate). Also, the very mechanism of definition/combination of concepts is stressed (logic of terms vs. logic of assertions). In these tentative theories, *paradoxes play the role of experimenta crucis*. This line of thought gives impulse to the elaboration of syntactical methods within combinatory logic, and to the rise of Recursion Theory.

"The inconsistency of certain formal logics": Kleene, Rosser and Curry

We consider two attempts: Church's approach, which substantially fails with unexpected positive consequences; Curry's, which gives birth to a long–running research program, eventually merging with the precious leftover (lambda calculus) of Church's system. Both systems of Curry and Church share a common destiny, that of being shown inconsistent by means of a non-trivial version of Richard's paradox. It was then Curry himself which discovered a simple, streamlined logical paradox pointing out a fundamental obstacle to the realization of the functional approach.

Since our main concern is not Curry's foundation of combinatory logic,[83] we only mention those points, which are relevant to the issue of paradoxes. Developing ideas of Schönfinkel [1924], and aiming at a mathematical clarification of the substitution process, Curry's 1930 thesis introduces a formal language based on basic combinators B (composition), C (permutation), W (duplication), K (cancellation),[84] Q (equality), and logical constants like universal quantifier, implication. Terms are then inductively generated by application from constants; intuitively, a term M stands for a function, and the application term MN denotes the value of the term obtained by replacing the first variable of M with N. Note that MM is admissible; so *direct self-reference* is not excluded from the system.[85] This feature tells us that the objects of combinatory logic cannot be simply interpreted as set-theoretic functions. Sometimes MN can be read as $N \in M$ and this is important for interpreting class theory and higher order logic.

The formal system comprises standard equations on combinators, rules for equality and the logical constants; its main goal is to derive equalities $X = Y$ and to make assertions of the form $\vdash X$. The idea is that combinatory logic has to be an "Urlogik', a theory which analyzes the modes of combinations of formal objects, substitution and the notions of proposition and propositional function: in a nutshell, it must give a foundation of all possible formal systems (of mathematics and logic). Among its specific subgoals, combinatory logic must eliminate paradoxes; this is achieved by turning paradoxes into theorems of pre-logic, which should be mathematically rendered by combinatory logic. For instance, Russell's contradic-

[83] For a good survey, see Seldin [1976], [1980].

[84] These constants satisfy: $Bxyz = x(yz)$, $Cxyz = xzy$, $Wxy = xyy$, $Kxy = x$.

[85] As already noticed, a much earlier instance of direct self-application, can be found in Hilbert's lectures of 1905, see [Kahle, 2004], cit.

tion amounts to the proof that a certain function F, defined by $F(\varphi) = \neg\varphi(\varphi)$, cannot be either true or false (see Behmann's paper [1931], p. 40).

Contrasting Russell's solution, Curry advances the idea that the root of paradoxes is not in admitting certain "combinations of concepts", but i*n admitting that these combinations are always propositions*. So the notion of proposition becomes a *theoretical concept*, that is decided by the theory and not simply stipulated in the metatheory. Types – as ranges of significance – are not assigned a priori to the expressions of the calculus; instead, they are *inferred by the calculus*, which assigns them – whenever it is possible – to its terms. These ideas contain *in nuce* fundamental developments (e.g. the *formulas-as-types interpretation*, see [Howard, 1980]).

For Curry – against Russell's prohibition, but also against Poincarè's views – it is legitimate to quantify over all propositions or all objects. The idea is that we understand a universal judgement in intension (i.e. relying upon general features pertaining to all possible objects of a given totality); we can make a general statement about all oranges, even if there is no definite collection of such fruits ([Curry, 1930]).[86] It is false to assume that a universal statement presupposes a definite, completed collection.

Unfortunately, appealing as they are, the ideas of a combinatory logic, as outlined in [Curry, 1930], [Curry, 1934b] had to be thoroughly revised. As we'll see, rejecting Russell's discipline was not without price.

A similar pattern is followed by Church in a different, independent style.[87] Church's logics ([1933], [1941]) are motivated by technical problems related to an attempt of avoiding free variables in formal logic and at the same time by dissatisfaction with Russell's and Zermelo's solutions that Church considers somewhat artificial. This leads him to conceive the formalism[88] of his theories as a general notation system for functions; Church's language is an applicative language, where there is only one basic category of terms (well-formed formulas in his terminology). Some terms are formally provable (or assertible) and are classified as true.

Terms are inductively defined from a set of basic constants and variables by means of application (denoted by $\{F\}(A)$, the result of applying F to A) and the characteristic lambda abstraction operator: if M is a term where the variable x occurs, $\lambda x.M$ is a term, naming the function defined by M.[89] Following Schönfinkel [1924], Church reduces functions of several variables to iterated application of functions of one variable; so $F(x, y)$ stands for $\{\{F\}(x)\}(y)$.

[86] For a related defense of impredicative definitions, see also Carnap [1932-33].

[87] Church does not cite Curry.

[88] Concerning Church's view, he has at this stage a conventionalist attitude, holding that there might be several possible systems, more or less convenient. He also strongly advocates an abstract axiomatic point of view: "the initial set of postulates must of themselves define the system as a formal structure, and in developing this formal structure reference to the proposed application must be irrelevant". Certain statements are reminiscent of Hilbert: In order to carry out the abstract treatment of logic, one must have an intuitive logic used for working with symbols (identificationy/distinction of discrete entities; Hilbert finitary mathematics).

[89] Note the restriction; so $\lambda x.c$ is not a term if c is a variable or a constant differing from x.

The basic constants designate logical operations: Π (a kind of restricted formal implication); Σ (existential quantifier), $\wedge$ (conjunction), $\neg$ (negation), ι (description operation) and A (generalized abstraction).[90] Without giving formal details, we recall that Church's system consists of (i) conversion (= computational) rules, which allow to replace terms with intensionally equivalent ones; (ii) rules for asserting certain terms as 'true'. In particular, there are:

1. "rules of procedures" (among them, the by-now standard rules of α– and β–conversion for λ–terms; Σ–introduction and Π– elimination;

2. several axioms (thirty seven in the first theory) ruling the remaining constants.

The new logic looks quite strong, since it can interpret naive class theory; for instance, A works like Frege's *Wertverlauf* operation and Church can define a class forming operation $K(M) = \{x|Mx\}$, deriving in [Church, 1933] naive comprehension (29-30, p. 856) and theorems 25-26, expressing the properties of the ϵ–operator in Frege's *Grundgesetze*.

Clearly the system is suspiciously strong and expressive. But Church's hope (or conjecture) is that contradictions be avoided by supplying logical axioms with formal restrictions ensuring the possibility that a propositional function be undefined for some arguments. Partiality is formally implemented by assuming that the functions involved in the logical axioms are non-empty (i.e. they truly apply at least to one object). So axiom 1 (reflexivity of Π) is stated in the form $\Sigma(F) \rightarrow \Pi(F, F)$.

Unfortunately, the project turned out to be unfeasible for unexpected reasons. After correcting the first version ([Church, 1932]),[91] a contradiction sprang out of an ingenious combination of lambda calculus with a form of the deduction theorem (see theorem I, p. 358, [Church, 1932]), which *does not require the axioms on A and class abstraction*:

> assume that M contains x free, $\Sigma(\lambda x.M)$ is provable and M is provable as a consequence of N, and x is free in N. Then $\Pi(\lambda x.M, \lambda x.N)$ is provable. If N does not contain x free, then N itself is provable.

Church was aware of the problematic aspects of his work and he actually set the agenda for his outstanding students in his address [Church, 1934] on the Richard

[90] $\Sigma(F)$ means that there is at least a value of x for which $F(x)$ is true; $\iota(F)$ means the object x such that $\{F\}(x)$ is true . As to $\Pi(M, N)$, it says that $N(x)$ is true for all values of x for which $M(x)$ is true; once we know that $M(C)$ and $\Pi(M, N)$ are true, then $N(C)$ is true (but we are not allowed to infer the truth of the implication $M(C) \rightarrow N(C)$). As to A, if F is an equivalence relation, $A(F, M)$ is "what M has in common with any N F–equivalent to M"; e.g. it satisfies in our language if F is formal logical equivalence:

$$\forall x(Mx \leftrightarrow Nx) \rightarrow A(F, M) = A(F, N)$$

[91] Inconsistent by a form of Russell's contradiction found by Church himself.

paradox, [92] where he discussed the adequacy problem of formal systems.

For him, a formal system could be termed adequate only if every *intuitively definable* number theoretic function F could be defined by a formula of the formal system itself; and *he believed that such systems could exist* ([Church, 1934, p. 358]), a kind of "Church's thesis" for definability. He even considered a more stringent requirement: a truly adequate system *must prove* the formal theorem $F : N \to N$, i.e. F should be *provably total* (or well-defined).

Then – he noticed – a supposedly adequate system falls prey of unpleasant consequences. Indeed, by exploiting Gödelian techniques (enumeration of formal theorems in the system itself) and its definitional power, we can effectively enumerate the provably definable functions of the system. But the enumeration function is intuitively definable; so, if it is provably definable, we apparently obtain a form of Richard's paradox. From this argument Church informally derives a Π^0_2-version of the incompleteness theorem: for every formal system, either the system is inadequate, as a tool for proving intuitively true statements of the form $F : N \to N$, or else the system proves theorems which are intuitively false. Roughly, an adequate system must be unsound. If this is the case, the verdict is highly negative; "it plainly implies that the whole program of the mathematical logician is futile" ([Church, 1934], p. 359).

The conclusions of the paper are rather obscure and speculative: according to him, "a way out of this condition of nihilism" (p. 360) is the search for a system of symbolic logic, where *the notion of implication is obtained by definition and there are a variety of implications obtained by different definitions.*[93] So, at this stage Church still believes that it is possible to escape the consequences of Richard's paradox and Gödel by a radical rephrasement of the logical theory of implication, insofar as it forbids an effective enumeration of the theorems of the form $F : N \to N$ within the system itself.

Be that as it may, less than one year later [94] Kleene and Rosser proved that Church's and Curry's systems can *provably enumerate* their own provably total definable number theoretic functions. Since the underlying theory of implication is standard in either formal system, they must be both unsound, as Church's version of incompleteness forecasts.

The result essentially requires (i) λ-calculus (rules I-V of Church) (ii) the logical axioms required to prove the deduction theorem I (cf. above). The non-trivial technical tools were already available in Kleene's development of number theory by means of λI-calculus (Kleene [1934], [1935]).

[Kleene and Rosser, 1935] is not self-contained and somewhat entangled, and it

[92] Delivered at a meeting of the Mathematical Association, Dec.30, 1934. According to him, the Richard paradox can be said to consist in the following problem. "How is it possible that a system of symbolic logic, in which the set of all formulas is enumerable should be adequate of any branch of mathematics which deals with the members of a non-enumerable set (...) ?", p. 357

[93] See Church [1941]; also Curry[1942b] and, for a recent implementation of these ideas, Myhill [1984].

[94] The paper [Kleene and Rosser, 1935] is received on November 13, 1934.

is difficult to appreciate its general import, which was neatly discussed in Curry's 1941 essay ([Curry, 1941]).

On the conceptual side, Curry distinguishes two completeness notions a formal system can enjoy. A system is *deductively complete*, if, whenever it derives a proposition B from the hypothesis A, then it also derives the implication $A \to B$ (deduction theorem or introduction rule for implication). A system is *combinatorially complete*[95] if, whenever M is a term of the system possibly containing an indeterminate x, there exists a term (Church's $\lambda x.M$) naming the function of x defined by M.

Curry's diagnosis is that the paradox of Kleene-Rosser arises because Church's and Curry's systems enjoy both kinds of completeness; hence it shows that the two properties are incompatible. On the technical side, Curry's system $\mathcal{S}$ axiomatizes the main ingredients exploited by Kleene and Rosser (with simplifications due to the fact that he works with λK-calculus). Besides the combinatory constants, he assumes only two terms N and F, representing the notions of being a natural number and the function space constructor. Actually, N can be dispensed with (cf. footnote 34, [Curry, 1941], p. 468) and F ensures that one can define a standard notion of implication; an induction rule on N is available for arbitrary terms of the system.

He then shows that (i) $\mathcal{S}$ can develop recursive arithmetic; (ii) there exists an enumerator, i.e. a term T such that, if a is the gödel-number of a closed term M and Z_a is the term formally representing a, then $TZ_a = M$ is provable in $\mathcal{S}$; (ii) there is a term Θ representing a primitive recursive enumeration of the $\mathcal{S}$-theorem such that

$$\mathcal{S} \vdash x \in N \to T(\Theta(x)).$$

So one can diagonalize out finally reaching a contradiction.

If one looks closely to Curry's proof, there is something unsatisfactory: while deductive and combinatorial completeness do not mention number theoretic notions, the inconsistency proof heavily hinges upon number theoretic construction. But, in a sense, number theoretic notions should be dispensable (they serve simply as an outcome of the underlying lambda calculus with the inferential power of standard implication).

That the detour through number theory and gödelization, is unnecessary was discovered by Curry himself, when [Curry, 1941] was to appear. Curry presented a new and simple paradox in a paper with the same title of [Kleene and Rosser, 1935], "in deference to the original discoverers of the contradiction".[96] The main result is the following theorem:

> (1) assume that we have a combinatorially complete system, i.e. essentially a system of combinatory logic which comprises the standard property of equality and the basic axioms ensuring the definability of

[95] See p. 455, [Curry, 1941].

[96] Curry refers to Carnap [1934a]) for the central idea; see also [Hilbert and Bernays, 1939], vol.2, 254 and Curry's paper [Curry, 1934a].

Church's lambda with the corresponding beta conversion (B, C, I, W suffice);

(2) assume also that the system contains an implication operator $\supset$ satisfying, for arbitrary terms M, N

$$\begin{aligned} &\vdash \quad M \supset M \\ &\vdash \quad M \supset (M \supset N) \Rightarrow \quad \vdash M \supset N \\ &\vdash \quad M \quad \text{and} \quad \vdash M \supset N \Rightarrow \quad \vdash N. \end{aligned}$$

Then $\mathcal{S} \vdash M$, for every term M.

The proof is immediate with the assumptions of (2), as soon as we can construct, for any given term B a term A such that $A = A \supset B$.

It is at this stage that combinatorial completeness is essential. Two constructions are given. The first one is by direct self-reference and is regarded by Curry as a version of Russell's paradox (it is a special application of the first fixed point theorem of combinatory logic). Choose: $A = HH$, where $H = \lambda Y(N(YY))$ and $N = \lambda X(X \supset B)$.

The second construction is by indirect self-reference and it exploits the machinery of [Curry, 1941] and Kleene-Rosser. Informally, Curry relates it to Epimenides (it is an application of the second fixed point theorem of combinatory logic). If T is the enumerator, one sets: $U = \lambda X(TXX \supset B)$ and $A = UZ_u$, being u the Gödel-number of U.

It would be interesting to follow Curry's in his analysis of the solutions to the paradox, but this would lead into the realm of the theory of functionality, the history of combinatory logic and proof theory. It is enough to recall that according to Curry, a remedy would be to formulate within the system the very notion of proposition. This is a thorny route, and a way out typically leads to a hierarchy of canonical propositions (in some sense analogous to Church's levels of implication). As we'll see, this idea will have some revival in recent years, but Curry doubts that it is a natural style of doing mathematics and logic.

Criticizing contraction and negation

To the tradition of combinatory logic we owe two ways of solving paradoxes, which will be pursued at some length in the 1970's and the 1980's with quite different (and even diverging) motivations (think of [Grišin, 1974], [Grišin, 1981], [Gilmore, 1974], [Gilmore, 1980], [Scott, 1975], [Feferman, 1974a], [Feferman, 1975], [Aczel, 1977], [Aczel, 1980], [Myhill, 1984], [Flagg and Myhill, 1987b]).

In Curry's analysis there is a clear perception that the contraction combinator W, satisfying $Wfx = fxx$, is at the root of paradoxes. W allows to define paradoxical terms, like WWW; if N represents negation and $Babc = a(bc)$, then $W(BN)(W(BN))$ is a fixed point of negation, and it is a functional correspondent of Russell class. On the logical side, the functional character of W[97] justifies the

[97] W is already introduced in 1930-35, [Curry, 1942a], p. 60

inference

$$A \to (A \to B) \Rightarrow A \to B,$$

which is crucial for the derivation of Curry's paradox.

The critical role of the contraction operator W is explicitly noted by Fitch ([Fitch, 1936]).[98] In this paper one finds is a crystal clear diagnosis of Russell paradox as tracing back to contraction: you need to consider a function of *two* variables, diagonalize and regard such an object as a new *unary* propositional function, which only works if W is accepted.

The proposed logic is explicitly called "non-contractive"; it is in essence an illative combinatory logic where B, C[99] can be defined, but not W. Fitch also uses a universal quantifier which is sensible to the number of occurrences of the quantified variables (linear quantifier), so that

$$\forall x f(x), \quad \forall x f(x, x,) \ldots$$

are rendered with $G(f)$, $G(f, f)$,... However, the paper does not go beyond a succinct development of a fragment of classical logic with predicates, and a series of statements, e.g. a form of choice is provable in the system, but the system itself apparently fails to develop a theory of integers and real numbers.

A few years later Fitch [1942] laid down the base for a new approach to the problem of finding consistent combinatory logic systems, which was pursued in a long series of articles, spanning till 1980. [Fitch, 1942] and the subsequent contributions are not easy to understand, due to the heavy formal machinery, but Fitch's work contains *in nuce* important ideas, which will be essentially taken over and refined by the investigations of the seventies.

In present–day terminology, Fitch's contribution amounts to *the inductive construction of suitable term models, endowed with self-referential notions of class, membership and truth.* This means that the notion of truth and membership are simultaneously and constructively assigned by means of *positive clauses*, so that they give rise to monotone operators (mapping sets of expressions into sets of expressions).

This has the pleasant consequence of ensuring the existence of non-trivial consistent truth sets, either by a direct inductive argument or as a corollary of the Tarski fixed point theorem (about monotone operators in complete lattices).

As to the specific developments, at first (see [Fitch, 1942]), Fitch explored a purely positive combinatory system K[100] with the explicit aim of defining a sort of universal formal system, where *every system of logic* could be represented.

More concretely, Fitch's system K is presented by means of an inductively defined truth set, which is closed under natural clauses for introducing certain basic

[98] Partly an adaptation of his Ph.D. thesis at Yale in 1934, *A system of symbolic logic that avoids the paradoxes without a theory of types*; in the paper Fitch thanks Church and Rosser.

[99] C satisfies the equation $Cabc = acb$.

[100] The choice of combinatory logic is supported by the observation that every logical system can be built up using only one mode of construction, i.e. application.

combinators (like B and W[101]) logical combinators corresponding to existential quantification, disjunction and conjunction, literal equality and *transitive closure*. For instance, if $a \mapsto a^*$ is the transitive closure operation, K will satisfy the closure condition

$$f^*ab \in K \quad \text{if and only if} \quad fab \in K \vee \exists c((f^*ac) \in K \wedge (f^*cb) \in K)$$

It is also worth noticing that class abstraction $\{x|A\}$ becomes definable in K and it satisfies forms of *untyped comprehension* (p. 110, [Fitch, 1942]), e.g:

$$M \in \{x|A\} \in K \quad \text{if and only if} \quad A[x := M] \in K.$$

The approach is radically intensional: classes are always classes of expressions M in some language (say, of the basic logic) and they are identified with attributes, while membership is essentially reduced to truth in the sense of K. Thus, that $M \in \mathcal{T}$ holds, essentially means that M truly falls under the property specified (or expressed) by T, i.e. that we can prove $TM \in K$. Accordingly, a set $\mathcal{T}$ of K-expressions is *represented in* K if and only if, for some expression T,

$$M \in \mathcal{T} \Leftrightarrow (TM) \in K$$

The basic logic K – although negationless –is nevertheless non–trivial; it allows *direct self-reference*[102] and it can represent sets given by *finitary inductive definitions* (see p. 112, theorem 5.16). Thus K works as a sort of *universal metalanguage for dealing with arbitrary finitary systems* [103] *and recursively enumerable sets*, and it is undecidable. For instance, K can represent first order predicate logic, the ramified version of type theory in *Principia Mathematica*, or even K itself. In [Fitch, 1944] it was further established that the collection of classes of K-expressions which are represented in K coincides with the collection of recursively enumerable classes of K-expressions. A later simplification of basic logic is due to Fitch [1954b].[104] and Myhill [1952]

Bochvar's logical calculus: the interplay between internal and external logics

Bochvar [1937] bases his proposal on the introduction of a three valued logic: besides the standard truth values T (truth) and F (false), he considers a third value N, to be interpreted as 'meaningless'. Not surprisingly, his logical analysis leads to the conclusion that the paradoxes involve meaningless statements; however, the

[101] B is named o; the system also comprises additional combinators ϵ, ϵ' for permutations and a strengthened composition o', satisfying: $o'fgab \in K$ if and only if $f(ga)b \in K$. The constant combinator is not included.

[102] I.e. diagonalization tricks are not grounded in the formalization of syntax, but are built in the untyped applicative structure, which underlies any logical constructions.

[103] Curry's inspiration is quite active here. In the first footnote of [Fitch, 1942] the author also acknowledges his own debt to the papers of Church, Rosser, Quine, Carnap and Tarski.

[104] It is shown that transitive closure is eliminable via a fixed point construction.

result is obtained by means of an axiomatic system, which is based on a non-classical logical calculus, which is, nevertheless, able to interpret classical logic (compare with later solutions, e.g. [Feferman, 1984]).

Indeed, a characteristic feature of Bochvar's formalism is that it distinguishes two types of connectives, roughly corresponding to two different modes of assertion. A statement A in itself assumes exactly one among the prescribed values (true, false or meaningless); these are the *internal truth values*. But then the higher level statement "A is true" is true or false; whatever value A assumes, e.g. even if the truth value of A is N, "A is true" is false. Bochvar's idea is to equip the internal three-valued logical operations with external logical operations, which correspond to statements on the metalevel and allow to use classical logic in dealing with non-classical statements.

Formally, he describes by means of three valued truth tables the main propositional internal connectives $\cap$ (and), $\sim$ (negation), $\cup$ (or), $\supset$ (implication), $\supset\subset$ (logical equivalence). The truth values of $\sim A$, $A \cap B$, ... coincide with their classical values if A, B assume classical values; they become meaningless (assume value N), as soon as one among A, B has value N (strictness). In particular, no formula built up with the classical connectives can be valid (or a tautology, i.e. true under all possible assignments), as $A \to A$ has value N, if A is meaningless. So no formula is valid under this three valued semantics.

Then there are connectives allowing the formation of *metatheorical statements* $\vdash A$, $\neg A$, $\downarrow A$, $\bar{A}$, to be read as "A is true", "A is false", "A is meaningless", and "A is not true" (in the given order). The value of $\vdash A$ ($\neg A$) is true if A is true (false), false (true) otherwise. As to $\downarrow A$, it is true exactly when A is meaningless; and $\bar{A}$ is false if and only if A is true. So one can define external versions $\to$, $\vee$, $\wedge$ of $\supset$, $\cup$, $\cap$ and an equivalence $\equiv$, such that $A \equiv B$ means that A and B have the same truth value. As to their relation, one has, for instance, the following valid statements:

$$\begin{aligned}(A \supset B) &\to (A \to B)\\ (A \cup B) &\to (A \vee B)\\ A &\leftrightarrow (\vdash A)\\ (A \equiv\sim A) &\equiv\downarrow A\end{aligned}$$

The first two arrows cannot be reversed.

An immediate consequence is that in the new logic the notion of tautology becomes non-empty, e.g. trivially $A \to A$ is valid. Furthermore, the external logical operations make it feasible to project possibly meaningless statements, which are built up using the ordinary connectives, onto meaningful statements by means of the transformations:

$$\begin{aligned}(\sim p) &\mapsto \bar{p}\\ (p \supset q) &\mapsto (p \to q)\\ (p \cup q) &\mapsto (p \vee q);\end{aligned}$$

and this yields an embedding of classical tautologies into tautologies with the external language (see theorem VI, p. 95, [Bochvar, 1937]).

The next step is to introduce a statement calculus Σ where paradoxes can be analyzed. Essentially, he describes a version of the extended type free functional calculus of Hilbert-Ackermann (1928), where propositions and propositional functions can also be substituted for individual variables with a basic limitation. Bochvar's restricted substitution rule amounts, in our terms, to a comprehension schema of the form

$$\exists\varphi\forall x(\phi(x) \equiv \alpha(x))$$

where α contains only *internal* logical operations.

As to the underlying logic, the propositional basis of Σ is formed by Bochvar's tautologies with the detachment rule for $\rightarrow$ and an introduction rule for $\cap$; the quantifier axioms are standard, but a strictness axiom ensures that a propositional function is meaningless if and only if it is meaningless for at least one witness.

Bochvar's formalism has some obscurities; he introduces an additional sort of variables $x_c, y_c \ldots$, where the c–label means that they range over classical (i.e.internal) propositional functions and propositions and he requires that the *individual variables* in the substitution rule can be replaced only by classical propositions or functions. But the formal details are not fully spelled out.

The paradoxes are readily disposed of in Σ. If we choose $R(\varphi) :=\sim \varphi(\varphi)$, we are driven to

$$R(R) \equiv\sim R(R),$$

which yields only the provability in Σ of $\downarrow R(R)$ and $\downarrow\sim R(R)$. Observe that the restriction in the comprehension or substitution is actually essential: if we replace $\sim$ with the external negation in the Russell predicate R, we are led to an inconsistency (as $A \equiv \bar{A}$ becomes a two-valued function of A). Similar conclusion can be drawn from the Grelling paradox.

It should be apparent that Bochvar's solution cannot be simply classified as a gap solution, where the logic is weakened, in order to get rid of paradoxes; instead, his way out embodies the distinction between level and metalevel, which is basic to the Tarskian solution or even to the type theoretic routes, in the logic itself and this makes his framework quite expressive (e.g. it can handle the very notion of 'non-sensical','meaningless'). [105]

Non-terminating processes and cycles

Behmann's paper [1931][106] offers a diagnosis of the paradoxes, which ascribes the fault of the paradoxes to the definition machinery. The paper (read in 1929) begins with an analysis of Russell's paradox in the form of predicate application.

[105] C.f. the historical notes of Feferman [1984] Bochvar anticipates Halldén [1949]; related work is due to Segerberg and Woodruff.

[106] Behmann's work and his role in Göttingen's foundational debate – also on the basis of his unpublished work including his dissertation of 1918 – has been thoroughly investigated in [Mancosu, 1999] Here we only briefly touch on his published essays.

Behmann observes that, once we have introduced the Russell predicate F by the definition $F(\varphi) =_{\mathrm{df}} \neg\varphi(\varphi)$, it must be possible to eliminate F from any argument involving it. But, if we try to replace F by its definiens in $F(F) \equiv \neg F(F)$, we find ourselves entangled in an infinite regress, as there is no F–free expression that could replace $F(F)$. So *the contradiction clearly shows an error in the theory of definitions*, namely to the use of definitions that give rise to an infinite chain of substitutions, without converging to a result; according to him, neither the naive concept of set, nor the notion of infinite is responsible for the contradictions. For Behmann all problems involved in the axiom of reducibility, predicativity, constructivity can be avoided in solving the paradoxes. His hint is to reform logic without resorting to type theory, but by adding an operator !, which, when we are given a predicate χ, singles out exactly those arguments x to which χ meaningfully applies. ! controls universal quantification; we can pass from a universal statement $\forall x\varphi$ to an instance $\varphi[x := a]$ only if a makes sense. Behmann does not develop a systematic theory and it is also unclear how to interpret his special operator '!'. Only much later he will develop a formal theory in his [Behmann, 1959]. Nevertheless, his contribution is conceptually valuable and it has inspired recent work [Aczel and Feferman, 1980], since, as alternative route to the *Principia*, it underlines the intensional aspects (definitional equality, or equivalence) of the paradoxical arguments, e.g. the existence of possibly divergent computational processes.[107]

In quite different contexts, Lewis and Langford [1932] are driven to conclusions which are not dissimilar from those of Behmann.[108] Their diagnosis is that the paradoxes show that certain conceptual complexes do not express propositions (where, roughly, propositions corresponds to 'possible facts' which are susceptible of being realized). They introduce the special notation $p : \alpha$ to mean that p is a name, whose meaning is the proposition α (so p and "α" denote the same entity and can replace each other); typically the Liar amounts to 'p: p is false', but we can also imagine more complex self-referential situations:

$$\begin{array}{lcll} p_1 & : & p_2 & \text{is false;} \\ p_2 & : & p_1 & \text{is false;} \end{array}$$

or even

$$\begin{array}{lcll} p_1 & : & p_2 & \text{is false;} \\ p_2 & : & p_3 & \text{is false :} \\ p_3 & : & p_1 & \text{is false.} \end{array}$$

Consider the first case. If we assume, say, that p_1 is true, then p_2 is false; hence by replacing p_2 with its meaning we get that "p_1 is false" is false, i.e. p_1 is true, i.e.

[107] See also the distinction between "Aufgabe" (task) of a substitution and "Ergebnis" (result).
[108] They devote chapter XIII to the discussion of logical paradoxes in their approach to logic, which has the semantical notion of proposition at its core.

we find our initial assumption. Again this does not explicate the meaning of the proposition; so we are entangled into a new replacement, and eventually "into an infinite sequence of expressions which are more and more complicated". A similar situation arises if we assume that p_1 is false. It is clear that, unless the sequence is broken somewhere, the expression remains meaningless. So we do not have a contradiction, but a vicious regression (p. 440), hence no proposition.[109]

In general, the lesson is that one can create arbitrary complicated cycles and check that they can lead either to contradictions (see the second case) or to infinite regressions. In either case, the conceptual complex fails to converge to a definite proposition.

3.4 Gödel on paradoxes

Why then does any attention to paradoxes survive at all ? Isn't true that after Russell Zermelo and Tarski had eventually invented substantial theoretical means to get rid of the conceptual malaise embracing the notions of class, set, truth, definability? The reason lies probably in a persistent interest for alternative formal paradigms, due to the unsatisfactory outcomes of *Principia Mathematica* and in the problematic place self-reference keeps playing in mathematical logic.

More essentially and deeply, as also witnessed by Wang in his book [2001, ch. 8] and by Myhill's report in [1984], Gödel has a foremost, genuine interest for concept-theoretic (or intensional or property-theoretic) paradoxes, which he sharply distinguishes from semantical ones.[110] and which he considers unresolved by the standard Zermelian set theory

With this in mind, an illuminating analysis of the situation in the mid–forties is offered by the classical Gödelian essay of 1944. According to [Gödel, 1944], the most significant contribution of Russell to the analysis of the concepts of formal logic concerns the logical paradoxes and their solutions (p. 131). In this respect, it is especially urgent to settle the question of the vicious circle principle as a principal tool for solving the paradoxes. Gödel's conceptual analysis is penetrating and gets rid of many prejudices.

The vicious circle in the form "no totality can contain members definable in terms of this totality, or members involving or presupposing the totality of its values" is regarded as highly problematic by Gödel; for him, it comprises three different principles. The first one (mentioning definability) is only applicable if mathematical entities are regarded as constructed via a definition: only if a constructivistic position is assumed, impredicative definitions are not allowed. On the other hand, the remaining versions are not admissible. For instance, we can legitimately describe an object by referring to a totality to which the object belongs: this is a feature of a description, not of the object. It is also conceivable that there is a consistent notion of circular class (he here mentions classes involving themselves à la Mirimanoff), and self-reference in itself is not a problem: consider

[109] The situation is reminiscent of the idea of semantical groundedness, see 5.1.

[110] E.g. the Liar is semantical, but the predicate version of Russell's one is concept-theoretic.

the sentence: "every sentence contains at least one relation word". Clearly it is true and it also belongs to the same totality to which it refers.

Of course, in agreement with the criticisms of Ramsey, Chwistek and the results of Tarski, he praises the theory of simple types as a sufficient means for avoiding the paradoxes (not only logical, but also epistemological, see p. 135).

However, his insights are not limited to the realm of set theoretic foundations (or to the restricted type theoretic view), as one would expect from one of the canonical texts of Gödelian realism. Surprisingly, he supports ideas that go beyond the border of his supposedly well-established philosophical point of view. So he notices that the theory of types "brings in a new idea for the solution of paradoxes, especially suited to their intensional form. It consists in blaming the paradoxes not on the axiom that every propositional function defines a concept or a class, but on the assumption that every concept gives a meaningful proposition if asserted for any arbitrary or objects as arguments"([Gödel, 1944], p. 150).

The problem with type theory is that it implements an additional principle, i.e. the idea that objects belong to different mutually exclusive categories,[111] and this makes the expressive power of type theory limited (e.g. we can't meaningfully state the fact that a certain object is not of a given type). So Gödel suggests that there might be possible "to assume every concept to be significant everywhere except for certain "singular points" or "limiting points", so that the paradoxes would appear as something analogous to dividing by zero." The goal is to devise a system showing that our logical intuitions are substantially correct (up to minor corrections). So Gödel inclines towards ideas which are apparently quite close to those of Curry and Church.

This is a sort of research program in a nutshell, that it will be occasionally pursued by several logicians, but most of the time, in isolation and, to a certain extent, lying dormant. At this point Gödel warns that the supposed partiality is in fact essential: "it is not possible to extend a concept to all arguments by defining another one which yields a false proposition in case the original be meaningless"; for instance, it is not possible to make the concept "meaningfully applicable" total.[112] Here Gödel explicitly observes that there have been attempts in this direction by Church ([1932], [1933]), but we already know about their failure ([Kleene and Rosser, 1935]).

Gödel even considers a further alternative to the Russellian view: Quine's work ([Quine, 1937]) is recalled twice, both as a "stepping stone" for a more satisfactory type theory system and as a version of Russell's zig-zagness.[113] However, Gödel

[111]The Gödelian formulation is reminiscent of the principle leading to semantical categories in the sense of the Polish tradition (Ajdukiewicz) and Husserl.

[112]By a form of Grelling's paradox.

[113]Quine's theory amounts to restricting the naive comprehension schema to those instances which are *stratified*, i.e. that can be read as meaningful type-theoretically. More formally, a formula φ is *stratified* if and only if it is possible to assign a natural number (type in short) to each variable occurrence of φ in such a way that

1. if $t \in s$ is a subformula of φ, the type of s is one greater than the type of t;
2. if $t = s$, is a subformula of φ, then t, s, r receive the same type;

does not mention that paradoxes were still quite close to NF. In 1942 Rosser [114] had published a form of Burali-Forti's antinomy in a seemingly natural extension (named ML) of the system NF, extended with "ultimate"classes. ML was intended to amend certain weaknesses of NF (e.g. with respect to number theoretic induction).

Once more, the theorem brought about an unexpected output of a paradox on set theory and the foundations of mathematical logic.

4 NON-SET THEORETIC FOUNDATIONS AFTER 1945

For the sake of historical analysis, it is helpful to partition the period under scrutiny in two different intervals: 1945-1975 and 1975 till the end of the century. The reason is that they have *marked different features*.

As to the investigations covering approximately thirty years between 1945 and 1975, the search for type-free non-set-theoretic foundations remains somewhat marginal and still a matter of isolated research programs and traditions (Curry's school, Fitch), scarcely interacting with the mainstream of mathematical logic.

Nevertheless, there is a significant renaissance of semantical investigations, which crop up in connection with philosophical questions and the analysis of natural languages (the notion of presupposition, denotation of singular terms, paradoxes involving epistemic notions, refinements of old ones).[115] *Semantical and foundational investigations hardly meet together*, even if logical problems are essentially related.

In the second part, since (roughly) 1975, the 'Leitmotiv' is different: work arising from paradoxes becomes technically much sophisticated and closer to mainstream mathematical logic; there are interesting conceptual byproducts, e.g. in the discussion of paradoxality, intensionality or the philosophical question about "the bearer of the truth". The investigation of semantical paradoxes produces non-trivial flexible "semantical machines", which tend to overcome the borders of isolated traditions and stand out for themselves as mathematical tools.

At the same time, the study of type free comprehension and property theory receives a substantial impulse by parallel axiomatic and model theoretic investigations, and applications to foundational problems (constructive foundations above all). We also underline that in this period the semantical and the foundational traditions meet together (this is particularly evident in works due to Feferman and Aczel).

3. all free occurrences of the same variable in any subformula of φ have the same type;

4. if x is free in ψ and $\forall x\psi$ is a subformula of φ, then the 'x' in $\forall x$ and the free occurrences of x in ψ receive the same type.

[114] Rosser notes that the same result was obtained independently by Roger Lyndon, then a student of Quine.

[115] E.g.Geach [1955], Kaplan-Montague [1960].

For convenience, we postpone the outline of the history of the proper semantical investigations to §5; below we mainly deal with the uncertain paths of type-free foundations, till their merging with constructive foundations (see §4.4).

4.1 Proof-theoretic work: Ackermann and Schütte

In the line of Hilbert's formal axiomatic method, there are contributions to be mentioned, essentially due to W. Ackermann ([1941], [1950], [1952], [1958], [1961]), who reviewed systematically Church's attempts in the thirties, and Schütte ([1953], [1954], [1960]).

The basic idea underlying Ackermann's and Schütte's approach stems from Behmann and Church: if we like to maintain the full naive comprehension schema, we have to assume that predicates are not in general *everywhere defined*. This causes troubles, as we cannot rely upon classical logic: we cannot accept *tertium non datur* (insofar as it is not true in general that the application of a predicate to an object provides us with a statement which is true or false), but we cannot dwell upon a standard (positive) implication either, for a paradox will occur as well, as noticed by Curry [1942b].

The original twist here is to define a logical calculus to which Gentzen's cut elimination techniques apply. Since we cannot go into a separate reconstruction of the single contributions of Ackermann and Schütte, we only outline the main technical ideas with reference to [Ackermann, 1961], which is a sort of conclusive survey (at least, of the research due to Hilbert's school in the field).

Formally, the language includes: (i) positive operations $\vee$, $\wedge$, $\forall$, $\exists$ and negation $\neg$, which is only allowed in front of atomic formulas and extended to the full language via De Morgan's laws; (ii) predicate symbols, corresponding to predicate application $\{-\}(-)$(in each finite arity, also designated by simple juxtaposition), equality[116] and predicate abstraction; (iii) one sort of individual variables; (iv) possibly, a list of primitive predicate symbols (depending on the applications one has in mind).[117]

The powerful type-free character of the language is made apparent by two features: first of all, the abstraction constructor provides unrestricted generation of new predicates, i.e. $\lambda x_1 \dots x_n.A(x_1, \dots, x_n)$ is a predicate, if A is an arbitrary formula, with n free variables; furthermore, *predicates do belong to the category of individual terms*, besides variables and individual constants.

The underlying logical calculus Σ_1 includes: (i) standard introduction and elimination rules for positive logical operations and quantifiers;[118] (ii) introduction and

[116] So that atomic formulas will have the form $a\,t_1 \dots t_n$ and $t = s$.

[117] They are qualified as 'special' by Ackermann.

[118] Equivalently, the reader can figure out the corresponding standard rules for Gentzen's LK sequent calculus; a variant of these is indeed embodied in Schütte's version of the system in [Schütte, 1960].

elimination rules for λ-abstraction[119]

$$\frac{\mathcal{M} \vee A(\vec{a})}{\mathcal{M} \vee (\lambda \vec{x}.A(\vec{x}))(\vec{t})}; \quad \frac{\mathcal{M} \vee (\lambda \vec{x}.A(\vec{x}))(\vec{x})}{\mathcal{M} \vee A(\vec{t})}$$

Last but not least,[120] Σ_1 is an infinitary (though constructive) calculus, as it is enriched by a form of ω-rule, due to Lorenzen: if to each n-tuple $\vec{t}$ of closed term we can effectively assign a derivation of $A(\vec{t})$, then we can infer $A(\vec{x})$, with the free variables $\vec{x}$.

Crucial is then the choice of the axioms for Σ_1. If p is a primitive predicate, p must come equipped with an effective method which partitions the universe. So p is well-defined everywhere and $p\vec{t}$ (respectively $\neg p\vec{t}$) is an axiom, provided $p\vec{t}$ ($\neg p\vec{t}$) holds true of $\vec{t}$.[121] Thus the primitive predicates constitute the building blocks over which the abstraction process takes place, according to a(n implicitly) predicativistic point of view.

The problem is that the "typenfreie Logik" meets severe obstacles. First of all, it is unable to prove all propositional tautologies, e.g. Σ_1 cannot prove $A \vee \neg A$, for arbitrary A. [122] This is because we cannot have axioms of the form $r\vec{t} \vee \neg r\vec{t}$, for arbitrary r, $\vec{t}$. Moreover, the system is incapable of representing hypothetical arguments: we cannot infer $\neg A \vee C$, once we have derived C from the assumption A.

It should be mentioned that Ackermann's and Schütte's investigations for implementing efficient consistent type-free logic were directly motivated by concrete foundational interests, like the search for smooth formalizations of mathematics, e.g. elementary analysis.[123] So there were attempts to overcome the above limits. Typically, Ackermann [1950] was led to study an extension of the logic Σ_1 with a new operator B,[124] such that BA (roughly) means that A is derivable. The new logic Σ_2 is consistent and it tolerates a weak form of deduction theorem in the form:

$$\frac{A \vdash C}{\neg BA \vee BC}$$

An immediate consequence is that $BA \vee \neg BA$ is a theorem of Σ_2, since A is trivially derivable from A; but, in general, there ought to be the crucial restriction that the rules involved in justifying $A \vdash C$ are those of the primitive system Σ_1 and hence *do not involve* B.

Clearly, B-logic cannot yield a deductively closed system; instead, the extension procedure can be iterated by introducing a new operator B_1, reflecting derivability

[119] Below $\vec{x}$, $\vec{t}$ designate finite sequences of variables or terms.

[120] This contrasts with the type-free logic investigated in [Ackermann, 1952].

[121] Typically, if we interpret $=$ as literal equality, we can count $=$ as a special or ground predicate.

[122] This amounts to "Aussagenlogische Unvollständigkeit". In sequent calculus terminology, $A \Rightarrow A$ is not derivable in general.

[123] The interested reader ought to compare it with the work done earlier in [Hilbert, 1918b] and in the IV Supplement of the *Grundlagen der Mathematik* [Hilbert and Bernays, 1939]. This aim also explains why extensions with the description operator and Hilbert's ε-symbol are investigated and proven consistent in [Schütte, 1960].

[124] B for "beweisbar".

with the rules of Σ_2. Therefore, one simply obtains *a hierarchy a logical systems*, corresponding to a hierarchy of deducibility relations, but none of them can be deductively satisfactory, though it is possible to show that the formalism is adequate to represent fragments of elementary and ramified analysis. Similar ideas are being pursued at about the same time by Fitch and later by Myhill (see [Fitch, 1946], [Fitch, 1952] on hierarchy of negations, [Lorenzen and Myhill, 1959], [Myhill, 1975], [Myhill, 1984]).

From a modern standpoint, Ackermann's and Schütte's idea can be better rephrased in semantic terms: one begins with a term model $\mathcal{T}$, endowed with a basic structure (defined via the basic predicates); the logical calculus just specifies an inductive construction, extending $\mathcal{T}$ to a higher order structure with classes and a non-extensional membership relation (application).

That the inductive construction is non-trivial (and it actually yields a model) is then granted by proof-theoretic investigations, based on *ordinal analysis*. In 1952 Ackermann produces a complex consistency proof in the style of Gentzen's second consistency proof of Peano Arithmetic for a finitary version of Σ_1: ordinal notations are assigned to derivations, and an ordinal lowering "simplification" procedure is defined, in such a way that consistency is granted by transfinite induction.

Schütte's most mature presentation is to be found in the classical *Beweistheorie* [Schütte, 1960]: there the cut elimination procedure for infinitary derivations is rather neatly adapted to type-free logic and carried out by a *triple* transfinite induction argument.[125]

4.2 From Fitch's combinatory logic to inductive definitions

As we know from 3.3, Fitch [1942] had already proposed a consistent combinatory system – the so-called basic logic K – which hinged upon the idea of partiality, but was rather poor for foundational aims, as it missed universal quantification and negation.

In 1948 the basic logic approach was much improved by the extended basic logic K' of [Fitch, 1948]. There truth is also closed under *universal quantification*, and there is a critical reappraisal of negation. In particular, *truth and falsehood are defined simultaneously and inductively*: together with each single clause for introducing a logical operator, one has also the dual condition for the same negated logical operator.[126] The initial clauses only involve atomic facts about the ground structure; so one starts with basic atomic truths (true literal equalities) and basic atomic falsehoods (false literal equalities).

[125] Cf. [Schütte, 1960], pp. 232-236. This reflects the higher complexity of the logic; besides the usual logical complexity of cut formulas and some measure involving the derivation of the cut premises, one needs a parameter controlling the abstraction rules.

[126] E.g. K' satisfies, together with the standard recursive clause that $A \wedge B$ is true in K' if and only if A and B are true in K', also the dual condition: $\neg(A \wedge B)$ is true in K' if and only if $\neg A$ is true in K' or $\neg B$ is true in K'.

This extension is especially important, since it allows to speak of the complement of a class within the logic. However, one has to carefully distinguish the *internal* complement $\overline{C}$ of a class C, i.e. the set of all t such that $(\neg(Ct)) \in K'$ from the external complement $-C := \{a \mid (Ct) \notin K'\}$ of C. In general, the first is only an approximation of the second (in the sense that $\overline{C} \subseteq -C$). So it is natural to select the collection of those classes which are completely represented in the system K', for which $-C$ and $\overline{C}$ coincide. A class X of expressions is *completely represented* in K' if and only if, for some expression M, one has:

$$\begin{array}{lll} t \in X & \text{if and only if} & Mt \in K' \\ t \notin X & \text{if and only if} & (\neg(Mt)) \in K'. \end{array}$$

The meaning of Russell's paradox, or, on the same par, of the Liar paradox, is that the above distinction is essential: while K' itself is representable in K' and hence satisfies a kind of *semantical closure*, K' can't be completely represented into itself.

Since K' is also closed under untyped comprehension and universal quantification, K' has a considerable expressive power, but it cannot be properly considered constructive (in a finitary sense), as the clauses for introducing the universal quantifier and the negation of the existential obviously correspond to infinitary rules.[127] It is possible to show that natural numbers can be completely represented in K', and, to a certain extent, K' allows the development of a theory of real numbers.[128]

Of course, due to the presence of infinitary inferences, the consistency proof for K' (i.e. of the fact that, for no c, both $c \in K'$ and $\neg c \in K'$) is carried out by means a generalized inductive argument (see theorem 3.2, p. 98), which goes far beyond the methods of Hilbert metamathematics.

Fitch's logic was further investigated and extended, mainly with foundational motivations.[129] Myhill [1952] proposed a kind of finitary metalanguage over K', in order to show that certain terms are theorems of K', and to pursue a nominalistic foundation of mathematics,[130] where only quantification over expression is admissible.

The most interesting results for mathematical logic involve the notion of representability over K' and *its connection with the theory of inductive definitions and hyperarithmetic sets*, as they are developed in [Fitch, 1954a] and above all in the subsequent note [Lorenzen and Myhill, 1959] .

[Fitch, 1954a] proves the *inductive completeness* of the language of extended basic logic K': all notions definable in the language of K' by means of the same

[127] If A, E are the combinators representing the universal and existential quantifiers (respectively), K' satisfies the closure conditions: (i) $Aa \in K'$ if and only if $ac \in K'$ for every expression c; (ii) $(\neg Ea) \in K'$ if and only if $\neg(ac) \in K'$, for every expression c.

[128] Although the collection of reals is not completely representable in the system, the least upper bound principle holds for completely representable classes.

[129] For instance, see the paper on the Heine-Borel theorem, [Fitch, 1949]. In the same vein almost forty years later, see [Flagg and Myhill, 1987b].

[130] In the sense of Chwistek and Hepter [Chwistek, 1948].

inductive procedure defining K' itself, are themselves representable in K'. In Fitch's own terms, let T be a relation of classes of K'-expressions to K'-expressions, which is represented in K', i.e. assume that there is a term t such that, if c represents a class C in K', then

$$T(C, a) \text{ holds} \Leftrightarrow tca \in K'$$

for every expression a. Given T, we can then specify a production rule $[T]$, such that $T(X, a) \Rightarrow a \in X$. A class C (of expressions) is "definable by transfinite induction by means of the rule $[T]$" (i.e. inductively definable by $[T]$ in the standard recursion theoretic sense, see Moschovakis [1974]), provided C is the least class closed under $[T]$. The main theorem of p. 32 establishes, by means of a fixed point argument:

(*) *the inductively definable sets of K'-expressions are exactly those representable in K'.*

[Fitch, 1954a] leaves open the problem of *dual representability*: if a class is represented in K' together with its complement, is this very class completely represented in K'?

In 1959 Lorenzen and Myhill give a positive answer, as a corollary of a thorough investigation of the constructive nature of inductively defined sets of numbers. They define the notion of *positive-inductive set* (of natural numbers). A set α of numbers is positive inductive, provided, for some list of parameters $t_1, \ldots, t_n$,

$$\alpha = \{n \mid (n, t_1, \ldots, t_n) \in I_\sigma T(\sigma)\}$$

where $I_\sigma T(\sigma)$ is the least set X satisfying

$$\{(n, t_1, \ldots t_n) \mid A(X, n, t_1, \ldots t_n)\} \subseteq X$$

and $A(X, n, t_1, \ldots t_n)$ is built up by means of $\wedge$, $\vee$, $\forall$, $\exists$ from atomic formulas of the form $t \in X$ and $r = s$ (r, s arithmetical terms; $(m_1, \ldots, m_k)$ is a primitive recursive function encoding finite number sequences).

It is then shown by means of a classical bar-induction argument:

(**) every Π^1_1 set is also positive inductive (Theorem 1) and hence a set α is hyperarithmetical if and only if both α and its complement α' are positive inductive.[131]

[131] This result is proved by Kleene [1955] and also in Spector [1961].An earlier step, hinting implicitly at the potential usefulness of recursion theory in semantic investigations, is due to the short paper Myhill [1950]. There he directly addresses the question whether any system, even without negation, can define its own truth. He gives a positive answer to the problem on the basis of standard recursion theoretic facts (existence of a total recursive function U enumerating the binary primitive recursive functions; representability of every r.e. set as image of a primitive recursive function). A much later trace in the same direction is to be found in exercise 16-91, p. 454 of Hartley Rogers's influential textbook [Rogers, 1967], where hyperarithmetical theory is applied as a means for producing a semantical characterization of HYP.

The interesting point is that [Lorenzen and Myhill, 1959] proves: (1) a set of U-expressions is represented in K' if and only if the corresponding set of Gödel-numbers is positive inductively definable (theorem 2); (2) if the set γ of all Gödel-numbers of the U-expressions belonging to a set B is hyperarithmetical, then B is completely represented in K' (theorem 3).

As a corollary, [Lorenzen and Myhill, 1959] yields a uniform representability theorem:

(***) there exists an expression L of K' such that, if m, n represent the class C and its complement in K' (respectively), then Lmn completely represents C in K'.

There are further original ideas in the paper: in particular, Lorenzen and Myhill consider a version of the hyperjump hierarchy[132] up to ω and they define a corresponding hierarchy $\{K'_i \mid i \in \omega\}$ of Fitch-like-systems. If one adds to the language an infinite sequence of new constants $\{U_i \mid i \in \omega\}$, it is possible to define

$$\begin{aligned} K'_0 &= K'; \\ K_{i+1} &= K'_i + \{U_i t \mid t \in K'_i\} + \{\neg(U_i t) \mid t \notin K'_i\} \end{aligned}$$

The hierarchy is suggestive of later formal developments (like iterated admissibility, universes in predicative type theories and over Frege Structures, levels of truth ([Martin-Löf, 1975], [Jäger *et al.*, 2001] [Kahle, 2002], [Cantini, 1995]).

In a series of subsequent papers (from 1958 to 1980) Fitch's systems are further enriched, but, as a matter of fact, they remain somewhat marginal and unrelated to other works. The main technical developments concern extensions with weak forms of extensionality [Fitch, 1958], [Fitch, 1963], [Fitch, 1980] and with modalities, needed for a formal clarification of the idea of l̀ogical essence ́(see [Fitch, 1967a], [Fitch, 1967b]). For instance, [Fitch, 1963] adopts as combinatory basis the standard combinators K and S and identity $=$ is no more interpreted as literal identity, but as denoting a form of generalized logical equivalence, which possibly tolerates the existence of propositions equivalent to their denials (cf. [Fitch, 1964], p. 399).

Modal logic opens up the route towards a consistent refinement of the naive comprehension principle in the form

(1) $\exists y \forall x (x \in y \leftrightarrow \Box \varphi(x))$

(φ being arbitrary, y not in ϕ). The Russell paradox then implies that there exists a set R such that $\neg R \in R$ and $\neg\Box\neg R \in R$, i.e. R is not a member of itself, but, possibly, it is a member of itself. Of course, Fitch is aware that inconsistencies can arise as well, if the properties of $\Box$ are not carefully chosen; indeed, the modal

[132] See for the main definitions, [Rogers, 1967], [Odifreddi, 1989].

logic he proposes cannot be *normal* (i.e. not closed outright under the box-rule: from φ, infer $\Box\varphi$; see [Fitch, 1967b], p. 107).[133]

Let us now try to assess the main contributions of Fitch's work. It is fair to say that he systematically exploits *the methods of the theory of inductive definitions and transfinite induction arguments* in order to prove consistency.[134]; in connection with his work, it first occurs that there is a link between *semantically closed languages and (generalized) recursion theory*

On the other hand, Fitch struggles to outline a unified treatment of semantics and abstraction, which to a certain extent stretches the boundaries of formal methods beyond the limits set out by Tarski's theorem. Paradoxes are ascribed to a vicious dependence of a sentence upon its denial: "The Russell paradox is the Epimenides paradox in its purest form, since we are provided with a statement which, on purely logical grounds, is equivalent to its own denial" ([Fitch, 1980], p. 32).

In particular, he shows (and he is aware of) that, once bivalence is weakened, *there exist semantically closed systems*, like his extended basic logic or the systems of [Fitch, 1963], where the identity combinator I provides for free a representation of the class of true propositions (cf. [Fitch, 1964], p. 400).

In a sense, the gap solution (see §5 below) is already in his papers with the main techniques (fixed points). But, surprisingly, papers in the same field hardly mention Fitch. This is probably due to the fact that his contributions were perceived as relevant only to a formalistic foundational program[135] and to illative combinatory logic. Especially this last point made (and still makes) a comparison with standard axiomatic systems, based on elementary languages and recursively defined categories of 'term' and 'formula', rather uncomfortable in general.

4.3 Combinators, classes and truth

The combinatory logic approach receives a new twist after the discovery of models for lambda calculus and combinatory logic, due to Scott and Plotkin. This is especially clear from the paper [Scott, 1975], which is an attempt to close up the gap between (standard predicate) logic and combinatory logic, in the light of the novel model-theoretic results. The work, motivated by the opening question

($\diamond$) *what is the relation between class abstraction and λ-abstraction in models of the λ-calculus?*

[133]Were it normal, we could infer $\Box\neg R \in R$ since $\neg R \in R$ is provable, whence a contradiction. It is also worth mentioning that the semantical investigation of these modal systems implicitly contains the idea of *supervaluation* (see the definition of *supplement* of a class in [Fitch, 1967a], p. 99).

[134]Some constructions are indeed explicitly recast by Fitch in the form of applications of the fixed point theorem for monotone operators, e.g. cf. [Fitch, 1967a], pp. 99-101. He also regards this method only weakly or moderately impredicative and even intuitionistic (reporting a remark to him by Myhill, ibid. p. 94).

[135]According to [Fitch, 1964], the existence of semantically closed systems is vital for vindicating the adequacy of formal methods to philosophy, against the idea that Tarski's theorem would destroy them for the very fact that no adequate truth predicate is available.

yields a neat reinterpretation of the *proper structural mathematical content* of Fitch's program. First of all, given an arbitrary lambda model M, the standard syntax of elementary logic[136] is uniformly represented by objects of the model M, using in a crucial way the fact that M is a combinatory algebra. The syntax encoding via M-objects preserves the *functional* character of the syntactical objects involved: e.g. to each formula $\varphi(\vec{a})$ with n parameters in the list $\vec{a}$, there is associated an operation, represented by the term $\lambda\vec{x}.\dot{\varphi}(\vec{x})$, expressing an operation $M^n \to M$. Of course, this allows the maximum freedom in carrying out direct self-referential constructions.

Then M is expanded to $M^* := \langle M, T_M, F_M\rangle$ with a pair $\langle T_M, F_M\rangle$ of subsets of M, simultaneously generated by transfinite induction.[137] and playing the role of *truth* and *falsity* The pair $\langle T_M, F_M\rangle$ evaluates the objects representing logically complex formulas according to the standard truth functional clauses, but the main point is that negation is never applied to T_M, F_M (e.g an object represented by $\dot{\neg a}$ is true (false) if and only if a is false (true)). So, as in §5.1, the construction amounts to expanding the ground model with a three-valued truth predicate.

The expanded structure can be roughly regarded as a model for Fitch's extension of basic logic (see [Fitch, 1948], §4.2) and the underlying lambda model M makes sense of membership/predication, predicate abstraction and truth by defining :

$$a \in b \;:=\; ba \tag{2}$$

$$\{x \mid \varphi\} \;:=\; \lambda x.\dot{\varphi} \tag{3}$$

$$Ta \;:=\; (\lambda x.x)a \tag{4}$$

Type free comprehension and the naive T-schema are elegantly subsumed under lambda abstraction and trivially, for every $a \in M$,[138] one has:

$$\dot{a \in \{x \mid \varphi\}}_M \in T_M \;:\Leftrightarrow\; \dot{\varphi[x := a]}_M \in T_M; \tag{5}$$

$$\dot{T\varphi}_M \in T_M \;\Leftrightarrow\; \dot{\varphi}_M \in T_M; \tag{6}$$

$$\ldots \tag{7}$$

The paper also sketches a proof theory axiomatizing the consequence relation for three-valued logic,[139] and it is suggested [Scott, 1975, p. 20] that there might

[136] There is an exception due to the semantics of three-valued logic. The primitive logical operators include also the so-called *squadge connective*, which gives the common part of $\wedge$ and $\vee$ and allows to introduce a formula $\varphi \overset{\vee}{\wedge} \psi$, such that $\varphi \overset{\vee}{\wedge} \psi$ is true (false) if and only if φ and ψ is true (false). The squadge with $\vee$, $\wedge$ and $\neg$ yields a system of connectives which is *complete for all monotone three valued operators.*

[137] Like the least fixed point model in the sense of Kripke (see §5.1).

[138] We assume that the language contains individual constants for each element of the ground universe; if t is a closed term, let t_M be the value of t in the lambda model M.

[139] It is understood that a sequent $\Gamma \vdash \Delta$ holds if and only if, whenever all elements of Γ are true, then some element of Δ is true, and whenever all elements of Δ are false, then at least one element of Γ is false, too.

be interesting formal systems, formalizing the properties of truth modelled by the construction of M^*. Last but not least, connections with inductive definability, which were developed by (and were somewhat hidden in) the complex formalism of Fitch and [Lorenzen and Myhill, 1959], are made perspicuous.[140]

In the final part, there emerges an unexpected link among *type-free* foundations, *constructive type theory* in the sense of Martin-Löf [1975] and De Brujin, and Curry's idea of functionality. If one considers the collection V of all definite or total classes[141] over a given lambda model M, expanded as above, that is

$$V := \{a \mid \forall x(x \in a \vee \neg x \in a)\},$$

V represents a rich *universe of types*. In fact, it is closed under exponential $\rightarrow$, generalized products Π and sums Σ, which are among the basic constructors in Martin-Löf's type theory. On the same par, the constructors can be shown to verify the type assignments rules for Π, Σ, $\rightarrow$.

Since $V \in V$ does not hold and V itself cannot be definite, [Scott, 1975] sketches the definition of a hierarchy $\langle T_n, F_n\rangle$ of truth/falsity predicates, which induces a corresponding hierarchy $\langle V_n \mid n \in \omega\rangle$ of universes, satisfying

$$V := V_0 \in V_1 \in V_2 \in V_3 \ldots$$

and closed under stronger and stronger recursion and induction principles. The construction parallels those of §4.2; it is analogous to the hierarchy of universes appearing not only in later developments of type theories, but also of Feferman's Explicit Mathematics.

Frege structures

A few years later there appears a work, related to the foundations of constructive type theory, due to Aczel [1977]; there, the author defines a *classical* interpretation of Martin-Löf's type theory with one universe [Martin-Löf, 1975], in a framework, which is amenable to a proof-theoretic analysis. Roughly, the formal system might be considered a counterpart to the step $M \mapsto M^*$ with a constructive flavour.

First of all, Aczel defines a system CFA of combinatory arithmetic, i.e. combinatory logic embedded in classical first order logic, extended with a predicate for natural numbers, which satisfies Peano axioms and is equivalent to Peano arithmetic. Then CFA is extended to a new system CFA_1 supplied with a notion of '*reflection* on CFA', i.e. with two predicates P and T, which classify the elements of the underlying combinatory algebra in *propositions* and *truths*. As to the axioms on P and T, they state:

[140] To each positive elementary operator form $\varphi(x, P)$ in the language of lambda calculus expanded with a truth predicate, there is associated an operation in the model represented by the lambda term $\lambda p\lambda x.\dot{\varphi}(x, p)$, such that the *set theoretic least fixed point* of the operator associated with $\varphi(x, P)$ is represented by the *combinatory fixed point* of $\lambda p\lambda x.\dot{\varphi}(x, p)$

[141] Of course, that a class b is total means that it splits the universe in two disjoint and exhaustive subsets, i.e. that for every a in M, either ba is true or ba is false.

- if $\dot{\varphi}$ is the form $\dot{\overbrace{x = y}}$ or $\dot{\overbrace{N(x)}}$, $P(\dot{\varphi}) \wedge (T\dot{\varphi} \leftrightarrow \varphi)$;
- $P(x) \wedge P(y) \rightarrow [P(\dot{\overbrace{x \rightarrow y}}) \wedge (T(\dot{\overbrace{x \rightarrow y}}) \leftrightarrow T(x) \rightarrow T(y))]$;
- $\forall x P(ax) \rightarrow (P(\dot{\forall} a) \wedge (T(\dot{\forall} a) \leftrightarrow \forall x T(ax)))$.

($\dot{\forall}$, $\dot{\overbrace{x \rightarrow y}}$ are closed terms representing the corresponding logical operations). Using the closure properties of P and T , it is shown how to interpret (a formulation of) the predicative constructive type theory ML_1 with a universe V (i.e. a type of small types) in CFA_1. The basic idea is that a is a *type* if and only if a is a *propositional function*, i.e. it satisfies $\forall x P(ax)$, while the judgement $z : a$ is simply translated with $T(ax)$ and *the type of small types* is given by the collection of all propositional functions.

As to the proof theoretic analysis, the strength of CFA_1 results from a sequence of intermediate steps, which involve the formal systems IAAC, ID_1^- and $\Sigma^1_1 - AC$. All systems have the same proof-theoretic strength, corresponding to *ramified analysis up to any level* $< \varepsilon_0$. IAAC is a subsystem of intuitionistic arithmetical analysis with choice, that it is interpretable in ML_1 and hence it yields the lower bound on ML_1.

As to the upper bound, the main idea is to exploit a formal counterpart of Scott's construction of M^*, since M^* is a model of CFA_1 (and hence validates ML_1). This can be done within the theory ID_1^-, which is simply the extension of Peano arithmetic with the axioms granting that each positive elementary operator form (in the arithmetical language; see [Moschovakis, 1974]) has a fixed point (in fact, it is not necessary to choose the least one, in order to validate CFA_1). Then ID_1^- is elegantly interpreted in $\Sigma^1_1 - AC$,[142] which is reducible by a result of Friedman to *ramified analysis up to any level* $< \varepsilon_0$. It is worth mentioning that some of the ideas in Aczel's paper have been later improved by Feferman [1982], in order to prove *the so-called Hancock's conjecture*, i.e. the proof theoretic equivalence between predicative analysis and the constructive type theory $\mathrm{ML}_{<\omega}$ with universes V_n, for arbitrary $n \in \omega$.

Along the same line of thought, Aczel [1980] introduces the notion of *Frege structure*, as a coherent context for the rigorous development of Frege's *logical notion of set* and an explanation of Russell's paradox, but also as "the structural counterpart of Martin-Löf's language for the foundations of constructive mathematics".[143]

In order to capture the sharp distinction between *objects* and *functions*, which is fundamental in Frege's thought, [Aczel, 1980] deals with a kind of second order variant of lambda models: the notion of *lambda structure*, which consists of

[142] $\Sigma^1_1 - AC$ is the subsystem of second order arithmetic, whose main schema is the choice axiom $(\forall x)(\exists Y)A(x, Y) \rightarrow (\exists Z)(\forall x)A(x, Z_x)$, where A is a first order formula. So no set quantifier occurs bound in A; $u \in Z_x$ shortens $p(x, u) \in Z$, p being a primitive recursive pairing function.

[143] See also the paper of J. Smith, [1984].

(i) a family $\mathcal{F}= \mathcal{F}_0, \mathcal{F}_1, \ldots$, where $\mathcal{F}_0$ is a collection of objects, and, for $n > 0$, each $\mathcal{F}_n$ is a collection of n-ary functions with values in $\mathcal{F}_0$, which is closed under explicit definitions;

(ii) a lambda system for $\mathcal{F}$, i.e. a pair of functionals $\lambda : \mathcal{F}_1 \to \mathcal{F}_0$ and $APP : \mathcal{F}_0 \times \mathcal{F}_0 \to \mathcal{F}_0$ such that, for all $f \in \mathcal{F}_1$, $a \in \mathcal{F}_0$,

$$APP(\lambda(f), a) = f(a)$$

A *Frege structure* has the form $F := \langle \mathcal{F}, \lambda, APP, P_M, T_M \rangle$, where:

1. $\langle \mathcal{F}, \lambda, APP \rangle$ is a lambda structure;
2. P_M, T_M are two subsets of $\mathcal{F}_0$ with $T_M \subseteq P_M$ (the collections of propositions and truths);
3. $\mathcal{F}$ includes a distinguished system of standard *logical operations*, under which the set P_M of propositional objects is closed and whose truth conditions are standard and specified by means of T_M. By contrast with [Aczel, 1977], only *the interpretation of implication is non-standard* (see Curry), as satisfies the axiom:

$$P(x) \wedge (T(x) \to P(y)) \to [P(\overbrace{x \to y}^{\cdot}) \wedge (T(\overbrace{x \to y}^{\cdot}) \leftrightarrow (T(x) \to T(y)))]$$

Of course, this makes *propositions dependent upon truth* and the existence proof for Frege structures relies upon a tricky application of the fixed point theorem for monotone operators.

Concerning the foundational aims, a Frege structure F shows that the logical notion of set can be made precise: if f is a propositional function, the *Fregean Wertverlauf* of f is simply the object of $\mathcal{F}_0$ represented by $\lambda(f)$, which, as in (5), plays the role of a class abstraction operation $\{x \mid f(x)\}$ and satisfies the naive comprehension schema. Of course, no contradiction arises, since by the Russellian paradox $\neg x \in x$ does not define a propositional function of F. The root of the contradictions are to be found, in Aczel's view, in the fact that Frege's *Grundgesetze* have exactly two propositions, the True and the False. But this would have the effect of rendering truth internally definable, against Tarski's theorem.

As to the logicistic program, it cannot implemented that far in F. For instance, it is not possible to define a propositional function, whose extension is an isomorphic copy of the natural numbers, though there exist Frege structures where such standard copies of N exist. More generally, if A is a set of objects of F, A is *internally definable* if and only if there exists a propositional function f_A in $\mathcal{F}_1$ such that, for every object a, $a \in A$ if and only if $f_A(a)$ is in the truth set of F. It turns out that the collections of propositions, truth and sets in a Frege structure are *not* internally definable, as well as the collection of subsets of a given set. Also extensionality in the sense of the internal membership relation fails.

To sum up: contrary to their original inspiration, Frege structures reveal themselves more suitable for the predicative and constructive foundations than for the impredicative logicistic foundations.

Extending Frege structures: from Fitch's K' to κ-continuous lattices

Myhill's works [1975] [1984] can be explicitly understood as a continuation of Fitch's work of the fifties (§4.2). However, it is just from the seeds of [Myhill, 1984] and the foundational weakness of [Flagg and Myhill, 1987b] that the interaction with lambda calculus models is deeply exploited by [Flagg, 1987].

First, a conceptual remark. Myhill's work strongly vindicates, in agreement with Gödel, that there never were set-theoretic paradoxes (and so the Cantorian notion of set is untouched by Russell's contradiction). Paradoxes only affect those conceptions which are inspired by the Fregean logical notion of set (as primarily given by a property). With this in mind, Myhill is mainly interested in refining Fitch's systems and in finding links with present forms of constructive mathematics.

The main problem with Fitch's system K' is of course *implication* and the fact that, as soon as implication satisfies modus ponens and the deduction theorem, Curry's paradox arises. The idea of Myhill is roughly this: in a non-classical context, an implication $A \to B$ is reasonably identified with a deducibility relation $A \vdash_R B$ with respect to a given set of rules R. But if we allow the natural introduction rule

(8) $\quad A \vdash_R B \Rightarrow \vdash A \to B$

"we cannot without circularity include amongst those rules *the rules for the very implication which we are trying to define*".[144] This leads Myhill in [Myhill, 1975], [Myhill, 1984] and [Flagg and Myhill, 1987a] to introduce a hierarchy of implications $\to_0$, $\to_1$, $\to_2$, ..., where $A \to_0 B$ is true if and only if A is deducible from B *without using any sort of implication*, $A \to_1 B$ is true if and only if A is deducible from B without using any sort of implication, *except possibly* $\to_0$, and so on. As to the falsity conditions, which are necessary in the present context, the rule is that $A \to_i B$ is false if and only if A is true and B is false.

As to the technical benefit, the enriched logical machinery allows to quantify over collections that can be *only positively* represented in the system (but they are not total or classical). In particular, there exist exponentials of different levels: if a is not total, one can define, depending on the level $i \in \omega$

$$[a \to_i b] := \{f \mid (\forall x)(x \in a \to_i fx \in b)\}$$

On the formal side, the hierarchy of implications gives rise to a system Σ_∞ of *illative combinatory logic*,[145] which includes:

(i) combinatory logic with extensionality;

[144] See [Myhill, 1984, p. 138].

[145] So formulas are identified with certain terms, to which combinatory logic applies; think of them as their representatives in a combinatory algebra, cf. Scott [1975]. In particular, logical constants and predicate constants become individual constants naming operations in the ground algebra.

(ii) natural deduction rules for introducing and eliminating terms of the form $M \wedge N$, $\neg(M \wedge N)$, $\forall M$, $\neg\forall M$;

(iii) closure rules for a predicate N, representing the set of natural numbers and a corresponding N-induction rule;

(iv) additional rules, ensuring that the underlying logic is classical on terms representing propositions (i.e. terms A for which $A \vee \neg A$ is provable[146] and that expressions of the form $t = s$, $N(t)$ represent propositions, i.e. that N and $=$ are definite or total);

(v) introduction and elimination rule for $\rightarrow_i$, where $i = 0, 1, 2 \ldots$, where the introduction rule for $A \rightarrow_i B$ has to meet the proviso that in the deduction of B from A only rules for implications *with lower index* $j < i$ are applied.

There are two problems left.

The first one is metamathematical: is the new system [Flagg and Myhill, 1987b] with levels of implication consistent? The answer is positive, and the techniques developed in [Flagg and Myhill, 1987b] to prove the theorems are the most interesting part of the work, since they lead the authors to apply inductive definability in depth and give evidence for the close two-sided link between self-referential truth and generalized recursion theoretic themes (cf. §5.3). After Fitch's work and Scott [1975], for every combinatory algebra $\mathcal{A}$ there exists a *classical N-standard Frege* structure $\mathcal{F}(\mathcal{A})$, consisting of (i) a subset $N_{\mathcal{A}}$ of $\mathcal{A}$ isomorphic to the natural numbers; (ii) a subset $\mathcal{T}$, which is consistent, i.e. for no $a \in \mathcal{A}$, $a \in \mathcal{T}$ and $(\neg a) \in \mathcal{T}$, and satisfies the closure conditions:[147]

$$\begin{aligned}
(a = b) \in \mathcal{T} &\Leftrightarrow \mathcal{A} \models a = b;\\
\neg a = b \in \mathcal{T} &\Leftrightarrow \mathcal{A} \models \neg a = b;\\
(Na) \in \mathcal{T} &\Leftrightarrow a \in N_{\mathcal{A}};\\
(\neg Na) \in \mathcal{T} &\Leftrightarrow a \notin N_{\mathcal{A}};\\
(a \wedge b) \in \mathcal{T} &\Leftrightarrow a \in \mathcal{T} \text{ and } b \in \mathcal{T}\\
(\neg(a \wedge b)) \in \mathcal{T} &\Leftrightarrow (\neg a) \in \mathcal{T} \text{ or } (\neg b) \in \mathcal{T}\\
(\forall b) \in \mathcal{T} &\Leftrightarrow ba \in \mathcal{T}, \text{ for all } a \in \mathcal{A}\\
(\neg\forall b) \in \mathcal{T} &\Leftrightarrow ba \in \mathcal{T}, \text{ for some } a \in \mathcal{A}\\
(\neg\neg a) \in \mathcal{T} &\Leftrightarrow a \in \mathcal{T}
\end{aligned}$$

Then the results of Fitch (§4.2), [Lorenzen and Myhill, 1959], [Scott, 1975] are generalized. Recall that a subset X of $\mathcal{A}$ is *(completely) representable* in $\mathcal{A}$ if and

[146] These inference amount to rules for introducing and eliminating double negation and to the distributive law linking universal quantifier with disjunction, corresponding to the schema $(\forall x)(A \vee B(x)) \rightarrow A \vee (\forall x)B(x)$, where x is not free in A.

[147] For the sake of simplicity, we do not distinguish (a) between objects of the structure and their names in the language; (b) between application terms of the combinatory language and the corresponding objects in the ground algebra $\mathcal{A}$.

only if there exists some c in $\mathcal{A}$, such that X (the complement of X) has the form $\{a \in \mathcal{A} \mid ca \in \mathcal{T}\}$ ($\{a \in \mathcal{A} \mid (\neg ca) \in \mathcal{T}\}$).

Let $Rep(\mathcal{A})$) ($Comp(\mathcal{A})$) be the collection of (completely) representable sets of $\mathcal{A}$, while $Ind(\mathcal{A})$ ($Hyp(\mathcal{A})$) is the collection of inductive (hyperelementary) sets in the sense of Moschovakis [1974]. Then, [Flagg and Myhill, 1987b] proves:

($\flat$) $Rep(\mathcal{A}) = Ind(\mathcal{A})$

Furthermore, if $\mathcal{A}$ is an *effective combinatory* algebra, that is, $\mathcal{A}$ is sufficiently similar to a *closed term model* of combinatory logic[148], it also holds:

($\flat\flat$) $Comp(\mathcal{A}) = Hyp(\mathcal{A})$

The effectiveness of $\mathcal{A}$ is then exploited by Flagg and Myhill for lifting to the context of Frege structures the fundamental stage comparison lemma of inductive definability,[149] which yields the main lemma for the consistency of Σ_∞, i.e. the *dual representability* theorem:

($\sharp\sharp$) There exists an object L in $\mathcal{A}$ such that, whenever a, b represent disjoint subsets X and Y of $\mathcal{A}$, then

$$u \in X \Leftrightarrow (Lab)u \in \mathcal{T};$$
$$u \in Y \Leftrightarrow (\neg Lab)u \in \mathcal{T}$$

($\sharp\sharp$) is the key step for modelling the hierarchy of implications of Σ_∞.[150] Indeed, let I_0 be the set encoding the deducibility relation $A \vdash_0 B$ of level 0 (no implication involved) while J_0 encodes $\{\langle A, B\rangle \mid A \in \mathcal{T}$ and $(\neg B) \in \mathcal{T}\}$. Since they are disjoint, there is an object $\mathcal{I}_0$ representing I_0, while its negation represents J_0. So one can (roughly) interpret $\rightarrow_0$ by means of $\mathcal{I}_0$ and this yields a consistent deducibility $\vdash_1$ of level 1. By applying the same procedure to the pair encoding $\vdash_1$ and J_0, one obtains a consistent interpretation of $\rightarrow_1$, and so on.

The second problem is foundational: to what extent can we proceed to formalize analysis in Σ_∞? According to the sketches of [Flagg and Myhill, 1987b], the use of levels of implication yields a version of the Dedekind reals in the system, and it allows to state and prove in Σ_∞ results like Peano's theorem for solutions of

[148] That $\mathcal{A}$ is *effective* means that there are an object $E \in \mathcal{A}$, two recursive functions App, Num satisfying the conditions:

1. for all a in $\mathcal{A}$, there exists a natural number n (in the sense of $\mathcal{A}$) such that $En = a$;
2. for all $k \in N$, $E(Num(k)) = k$;
3. for all m, n in N, $E(App(m, n)) = (En)(Em)$.

[Cantini, 1993] shows that the assumption that $\mathcal{A}$ is effective, is superfluous for proving the definability results of [Flagg and Myhill, 1987b].

[149] See [Moschovakis, 1974] According to this result, the relation comparing the ordinal stages of the elements of the least truth set over a Frege structure is expressible within the language of the structure itself.

[150] A consequence of dual representability is that the squadge operator is definable in the full language with truth; see [Flagg and Myhill, 1987b].

ordinary differential equations or the Heine-Borel lemma, which are not available neither in Fitch's nor within the frame of Frege Structures. However, the system is not that manageable, due to the stratification of implications and deducibility.

It seems fair to say that Σ_∞ is unsatisfactory and this is witnessed by the switch towards different sorts of strengthening for the illative approach. In [Flagg and Myhill, 1987b] the question is raised of "powerful metatheorems relating Σ_∞ to well-known systems". As proven later, Σ_∞ has an inner model in a theory of the same proof theoretic strength as ramified analysis up to any level $\alpha < \epsilon_0$. Hence it is not an essential step even beyond CFA_1 and constructive type theory with one universe.

A highlight of the interaction between lambda calculus and class abstraction is the subsequent work of Myhill and his student Flagg [Flagg and Myhill, 1987a], [Flagg and Myhill, 1989], [Flagg, 1987]. The initial observation is that Scott's lattice theoretic model D_∞ is an injective T_0-space (with respect to its Scott topology[151]) and hence an arbitrary function from a discrete subspace can be extended to a continuous function on the whole space. If one builds a Frege structure on the top of D_∞, the choice schema

$$(9) \quad (\forall x \in N)(\exists y)\varphi(x,y) \to (\exists f)(\forall x \in N)\varphi(x, fx)$$

holds for *arbitrary formulas* φ, as the natural numbers N form a *discrete subspace* of D_∞. This is essentially sufficient to prove the consistency of a natural formal (illative combinatory logic) system FL for Frege structures extended with a comprehension rule CR for discrete classes.[152] Then it is easy to show that FL+CR is strong enough to model second order arithmetic with full second order comprehension, and that it can be further enlarged to a system EFL, which also embodies a well-ordering principle of the universe, additional choice principles and can develop a considerable amount of analysis.[153]

However, in order to get higher type choice and comprehension principles, it would be desirable that the collection of discrete subsets be closed under exponentials or even under generalized sums and products. But this is not true for D_∞, since the collection of objects representing number-theoretic operations in D_∞ is *not* discrete. [Flagg, 1987] remedies to this limitation by a clever extension of Scott's lattice theoretic machinery. If k is a regular infinite cardinal, it is possible to construct an injective κ-space[154] D_κ, where the collection $\mathcal{DIS}$ of discrete

[151] See [Scott, 1972, p. 99].

[152] Indeed the property of being discrete can be easily represented in the language of Frege structure: in the sense of [Flagg and Myhill, 1987a], a set X is (formally) discrete if and only if there exists a well defined total property representing X which satisfies the condition that the equality relation restricted to X has a characteristic operation f, i.e. whenever a, b are elements of X, then $fab = 0$, if $a = b$, else $fab = 1$. This definition allows to prove in FL that N is discrete.

[153] There is a classical reference here: the fourth supplement of Hilbert-Bernays [1939].

[154] This mean that (1) the topology is T_0 and the intersection of fewer than κ opens is open; (2) every continuous function f from a subspace X to D_κ extends to a continuous function on the whole space.

sets is closed under generalized sums Σ and products Π. D_κ immediately implies the soundness of the extension of the theory FL of Frege structures, which (i) includes a comprehension rule corresponding to (9) with N replaced by (an object representing) an *arbitrary discrete* subset A; (ii) proves that the finite types generated from N are all discrete. The resulting theory goes far beyond higher order arithmetic and hence is fully impredicative; so it has no problem in formalizing ordinary mathematical analysis.

In [Flagg and Myhill, 1989] it is further proved that, if k is strongly inaccessible, the resulting D_κ-model verifies the axioms EFL* of *a theory of Frege structures, in which ZFC itself can be embedded*. Roughly, this is achieved by postulating - in the style of illative combinatory logic and Frege structures - axioms ensuring that a well-ordering of the universe is available, and that this *well-ordering has a discrete initial segment of strongly inaccessible length* [Flagg and Myhill, 1989, p. 85]. The definitional strength of combinatory logic implies the existence of a copy of the cumulative hierarchy with closure properties, granting that standard set theoretic operations are well-defined.

It can also be consistently added that *discrete classes are closed under direct images*. As observed by H. Friedman, this follows from the existence of extensional combinatory algebras, in which every finite set is *separable*. We recall that, given a combinatory algebra $\mathcal{A} = \langle A, \cdot\, k, s\rangle$, $X \subseteq A$, X is *separable* if any function $f : X \to A$ is representable by an element $f^\sharp \in A$, i.e. there exists $f^\sharp \in A$, such that

(10) for all $a \in A$, $f(a) = f^\sharp \cdot a$

Plotkin [1995] has shown that this is indeed the case. For instance, there exist extensional combinatory algebras having the power of the continuum, in which *every countable subset is separable*.

REMARK 3. A new attempt towards an integrated foundation of mathematics, logic and computer science on the basis of the primitive notion of self-applicable operation, is recently implemented by the so-called *map theory* MT, due to Grue (see e.g. the work in cooperation with Berline [1997], where a comparison with the Flagg and Myhill framework is sketched.).

4.4 Non-extensional formal theories of classes and operations

Towards the end of the sixties and culminating - in parallel to formal semantics (cf. Kripke 5.1) and the combinatory logic approach - around 1975, there are investigations which contribute to put in a proper perspective the technical essence of the previous work and open novel directions for applications of (some form of) untyped comprehension.

Again, as with the combinatory logic approach, a common feature is that all the proposals *give up extensionality*, and membership $\in$ is properly interpreted as *predicate application* (predication). The traditional abstract notation for classes $\{x \mid \varphi\}$, which is usually eliminable in extensional contexts, must be considered an

independent logical ingredient in its own, and, as such, it is thoroughly investigated by Gilmore [1974]–[2001].

It is realized especially through the work of Feferman [1974a]–[1984] that in many contexts *it is essential to make the uniform dependence of a given class construct* $\{x \mid \varphi\}$ *on its free parameters explicit*; moreover, it is natural and useful to enrich the picture of the logical universe with a theory of operations, that regards the *external operation*

$$\vec{v} \mapsto \{x \mid \varphi(\vec{v})\}$$

as an *internal operation*. This is an important conceptual feature that leads to theories where classes and operations of a very general nature interact to a certain extent: *we can apply operations to classes, but we can also classify operations*. In a sense, we are close again to the original inspiration of combinatory logic, but without assuming its "ontological monism" (it is no more true that *everything is a function*, since functions and classes, though living in the same universe, obey to distinct principles).

Combining unrestricted abstraction with standard logic

A rather influential paper for the investigation of self-referential systems - and implicitly for the clarification of the set theoretic core of Fitch's work - is Gilmore's [1974] (published in 1974, but presented in 1967), which makes the set theoretic construction behind Fitch's extended basic logic fairly explicit and outlines a plain theory of *non-extensional partial classes*, based on first order classical logic with identity. A partial class P is extensionally identified by a pair (X, Y) of disjoint subsets (of a given universe V), where in general, $X \cup Y$ is only a subset of V. If $X \cup Y = V$, the class is total and Y coincides with the usual set theoretic complement of X. However, the basic choice of [Gilmore, 1974] is to stick to a *non-extensional* point of view and to work directly with pairs of attributes (not with their extensions).

Formally, Gilmore's theory PST' includes an equality predicate $=$, and two binary predicates $\in$ (membership), $\notin$ (anti-membership); so prime formulas have the form $t = s$, $t \in s$, $t \notin s$.

An essential feature of PST' is that it involves *an abstraction operator for producing explicit notations for partial classes*. Thus terms and formulas are simultaneously and inductively generated by means of standard logical operators and the following clauses:

- if t, s are terms, $t \in s$, $t \notin s$, $t = s$ are formulas;
- if ϕ, ψ are formulas, then $\{x \mid \phi(x), \psi(x)\}$ is a term whose free variables include all the free variables of ϕ, ψ, except x.

$\{x \mid \phi(x), \psi(x)\}$ represents a (possibly) partial property P, whose elements satisfy ϕ but not ψ, while the complement is the set of all elements satisfying ψ, but not ϕ.

PST' comprises a consistency condition on membership and its dual, and two comprehension schemata ruling the abstraction operator:

1. $\forall x \forall y. \neg(x \in y \wedge x \notin y)$;

2. $\forall v((v \in \{x \mid \phi(x), \psi(x)\} \vee (\phi(v) \wedge \psi(v))) \equiv \phi(v))$

3. $\forall v((v \notin \{x \mid \phi(x), \psi(x)\} \vee (\phi(v) \wedge \psi(v))) \equiv \psi(v))$

The meaning of the comprehension schemata is easily grasped through their consequences. In particular, PST' actually implies a more elegant abstraction principle, which relies upon the idea of positive counterpart ϕ^+ of an *arbitrary* formula ϕ.

Define $t \neq s := \exists z(t \in z \wedge s \notin z)$; then ϕ^+ is the unique formula which is obtained from the negation normal form of ϕ by replacing each occurrence of $\neg t \in s$, $\neg t \notin s$, $\neg t = s$ respectively by $t \notin s$, $t \in s$, $t \neq s$. Clearly ϕ^+ is positive; if we let $\phi^- = (\neg\phi)^+$, we then have by consistency:

$$\phi^+ \rightarrow \phi \rightarrow \neg\phi^-$$

If $\phi(x)$ is arbitrary and $\{x \mid \phi(x)\} := \{x \mid \phi^+, \phi^-\}$, [Gilmore, 1974] shows in PST', for arbitrary $\phi(x)$:

$$\forall v(v \in \{x \mid \phi(x)\} \quad \leftrightarrow \quad \phi^+(v)); \tag{11}$$

$$\forall v(v \notin \{x \mid \phi(x)\} \quad \leftrightarrow \quad \phi^-(v)) \tag{12}$$

Let us now briefly recall the idea behind the model construction. First of all, the model's universe is the set of all closed terms of PST'. In particular, a term $\{x \mid \phi(x, v), \psi(x, v)\}$ with the free variable v shown, is interpreted as the function assigning to every element a of D the closed term $\{x \mid \phi(x, a), \psi(x, a)\} \in D$.

In order to evaluate arbitrary sentences, the strategy is to specify a (consistent) set M of atomic sentences: for this naturally induces an interpretation of the primitive predicates of PST', once we identify the meaning of $\in$ ($\notin$, $=$ respectively) with the set of all ordered pairs $\langle t, s\rangle$ such that $(t \in s)$ $(t \notin s,\ t = s)$ is in M (respectively).

The construction is implemented by transfinite recursion on ordinals. First, one defines the set M_0 of all true prime formulas, which do not involve $\in$, $\notin$, i.e. of the form $t = s$. Since $=$ is interpreted as literal identity, it is enough to choose $M_0 :=$ the set of all identity sentences $t = t$, where $t \in D$. Assume that we have defined M_μ; then $M_{\mu+1}$ consists of:

- the sentences of M_μ;

- all sentences of the form $t \in \{x \mid \phi(x), \psi(x)\}$ and $t \notin \{x \mid \psi(x), \phi(x)\}$ provided $\phi(t) \wedge \neg\psi(t)$ holds in the interpretation induced by M_μ.

If we choose $M_\lambda = \bigcup\{M_\beta \mid \beta < \lambda\}$, when λ is a limit ordinal, we obtain a non-decreasing transfinite sequence $\{M_\alpha \mid \alpha \text{ ordinal }\}$.

With the crucial help of the assumption that ϕ, ψ are positive, Gilmore then proves that there exists a countable limit ordinal κ such that $M_\kappa = M_{\kappa+1}$ and M_κ is a model of PST'. The non-extensional interpretation is in some sense intrinsic: PST' provably refutes extensionality, by showing that there are at least two provably different objects representing the empty set. For the proof, it is *essential* that the comprehension schema is in explicit form; in fact, if (11) is stated as a pure existential principle for positive formulas and the term constructor $\varphi \mapsto \{x \mid \varphi\}$ is omitted, one obtains a system *consistent with extensionality*. This follows from results due to Malitz [1976] in his 1976 Ph.D.Thesis, Weydert [1988] and more recently to Forti-Hinnion [1989].

P. C. Gilmore has further pursued the investigation of type free predicative and impredicative systems as suitable frameworks for the foundations of mathematics and computer science [Gilmore, 1980], [Gilmore, 1986], [Gilmore, 2001]), which are formalized in Gentzen style and proved consistent by semantic cut elimination arguments. All systems have an underlying partial logic and in this sense they develop ideas we already saw in §4.1, §4.2. The problem with these theories is that they have only term models and in the impredicative case the constructions are rather involved.

"Towards useful type-free theories"

In 1975, in parallel with (but independently on) substantial contributions in the field of formal semantics (c.f. §5, §5.1, §5.2), there appear two papers by Feferman[155] [1974a] and [1975], which mark a definite shift towards type-free systems inspired *by specific foundational needs more than programmatic or ideological motivations*. While [Feferman, 1975] is concerned with the development of systems for representing Bishop's constructive mathematics and logical issues arising from theoretical computer science,[156] [Feferman, 1974a] considers relatively unrestricted comprehension principles which permit *self-application* and are motivated by the formalization of theories of global structures, as occur in categorical approaches to algebra, topology and the like. All type-free systems are based to a certain extent upon the notion of *self-applicable operation*. Since a substantial body of experience is extracted from (generalized) recursion theory, operations are conceived as *partial* and *non-extensional*, given by possibly abstract rules, which (i) may not be defined everywhere and yet have no a priori bound on their domain of application, and (ii) are always associated (and identified) with a specific presentation.

The first option considered in [Feferman, 1974a] is to extend any given theory T by means of a *comprehension principle $\widetilde{C}$ for arbitrary monotonic formulas* and a corresponding application relation. Thus the language of T is expanded to a

[155] Presented in 1974.

[156] This will become explicit in later work by Feferman and Jäger during the second part of the eighties.

language $\mathcal{L}(\simeq)$ with a ternary application predicate $\simeq$; the corresponding atomic formula, usually written as $xy \simeq z$, means that the operation represented by x applied to y yields z as value.

Incidentally, expressions of the form xy, the so-called *application terms* (a.t. in short), are not properly part of the language, since they can be undefined. If t is a. t. and defined, we write $t\downarrow$. Application terms can easily be handled by natural contextual definitions,[157] or else one can adopt a logic of partial terms in the background (see [Beeson, 1985], [Troelstra and van Dalen, 1988]).

Of course, $\simeq$ satisfies a uniqueness condition $\widetilde{U}$:

(13) $xy \simeq z \wedge xy \simeq v \rightarrow z = v$

A formula φ of $\mathcal{L}(\simeq)$ (with parameters $\vec{v}$) is $\simeq$-*monotonic* if and only if , whenever $\mathcal{M}$ is a model for T, R, S are ternary relations on M (the domain of $\mathcal{M}$) interpreting $\simeq$ and $\vec{a}$ is an arbitrary assignment of values in M to $\vec{v}$, then

$$R \subseteq S \text{ and } \langle \mathcal{M}, R\rangle \models \varphi[\vec{a}] \;\Rightarrow \langle \mathcal{M}, S\rangle \models \varphi[\vec{a}]$$

If $\varphi(x, y, \vec{v})$ is an arbitrary $\simeq$-monotonic formula, the φ-instance of $\widetilde{C}$ is

$$(\forall \vec{u})(\exists f)(F_\varphi \vec{u} \simeq f \wedge (\forall x)((\exists! y)(\varphi(x, y, \vec{u}) \rightarrow fx\downarrow) \wedge \\ \wedge (\forall x)(\forall y)(fx \simeq y \rightarrow \varphi(x, y, \vec{u}))$$

Here F_φ represents the operation associating to $\vec{v}$ the unique y satisfying the formula $\varphi(x, y, \vec{v})$.[158] The main result is that

(α) every model of T can be expanded to a model of T+$\widetilde{U}$+$\widetilde{C}$, and hence:

(β) $T_\simeq := T + \widetilde{U} + \widetilde{C}$ is conservative over T.

For the proof, one recursively defines a non-decreasing sequence $\simeq_\delta$ of ternary relations, which are approximations of the proper application relation. The assumption on φ again ensures that the sequence converges to a relation which is closed under every instance of $\widetilde{C}$.[159]

---

[157] Here is the list of the inductive rules:

- if $t := x$, $t \simeq z$ is $x = z$;
- if $t = rs$, then $t \simeq z$ is $(\exists u)(\exists v)(r \simeq u \wedge s \simeq v \wedge uv \simeq z$;
- $t \simeq s := (\forall u)(t \simeq u \leftrightarrow s \simeq u)$;
- $t\downarrow := (\exists x)(t \simeq x)$.

[158] The definition of F_φ depends on the code of φ; it is always assumed that T contains a constant 0, a unary function symbol *suc* and a binary operation symbol P, with the axioms implying that P is an ordered pairing function, 0 with *suc* satisfy the standard postulates for zero and successor. This is sufficient for representing natural numbers in T and for encoding arbitrary formulas with parameters.

[159] The link with the theory of monotone inductive definitions is present to the author; see p. 77.

Consistent variations - again motivated by recursion theoretic principles - are possible (see e.g. the corresponding selection and reduction schemata). An important consequence of $\widetilde{C}$ is that the collection of operations represented by elements of M via application is indeed a *partial combinatory algebra* (so there exists K, S satisfying the $\simeq$-analogues of the usual combinatory equations), and hence a version of the second recursion theorem[160] is provable. It is also important to observe that the applicative structure is provably non-extensional

$$\neg(\forall f)(\forall g)((\forall x)(fx \simeq gx) \to f = g)$$

However, if classes are identified with two-valued operations, even if it is possible to have good closure conditions on classes, it is not enough for obtaining closure under an analog of the function space construction. This leads to consider alternative frameworks where *operations and classes are independently assumed as primitive notions.* Thus [Feferman, 1974a] contains a generalization of Gilmore's schemata, the so called "amplified η, $\bar{\eta}$-comprehension schema $\tilde{C}^{\eta}$: if φ_0, φ_1 are monotonic in η, $\bar{\eta}$, then there exists a term $\{x \mid \varphi_0(x), \varphi_1(x)\}$ such that

$$(\forall \vec{u})(\exists c)(\{x \mid \varphi_0(x, \vec{u}), \varphi_1(x, \vec{u})\} \simeq c \wedge (\forall v)(v \eta c \leftrightarrow \varphi_0(v, \vec{u})) \wedge \\ \wedge (\forall v)(v \bar{\eta} c \varphi_1(x, \vec{u})\} \leftrightarrow \varphi_1(v, \vec{u})))$$

(η, $\bar{\eta}$ stand for intensional membership and anti-membership, respectively). Since the defining conditions are allowed to contain the application relation, total classes turn out to closed under the operation $a, b \mapsto a^b$.

[Feferman, 1975] definitely sticks to a two-step strategy. If T is any first-order theory, T is enriched to an applicative theory $T_{\simeq}$, which embodies principles granting that the universe of partial operations is *a partial combinatory algebra*, expanded with definition by cases over the universe and an operation representing existential quantification on the universe. Then partial predicate application is introduced by means of *modally restricted naive comprehension principles* (see Fitch [1967a], section 4.2 with the schema (1), Lemmon [1963]). The logic is extended with an S4-modality operator $\Box$ and the comprehension schema assumes the form: if φ is an *arbitrary* formula of the expanded language, η, $\bar{\eta}$ stand for intensional membership and anti-membership, respectively,

$$(\exists f)(\forall \vec{a})(\exists c)(fa \simeq c \wedge (\forall x)(x \eta c \leftrightarrow \Box \varphi) \wedge \\ \wedge (\forall x)(x \bar{\eta} c \leftrightarrow \Box \neg \varphi))$$

The idea is that predication is given by a *generalized investigative procedure* $\mathcal{I}$, whose aim is to determine which elements of the domain fall under a given property.

[160] There exists a term R such that for every f, x,

$$Rf \downarrow \wedge f(Rf)x \simeq (Rf)x$$

$\Box\varphi$ informally means that the sentence can be verified (by means of $\mathcal{I}$) and the effect of the modality is just to make the verification process monotone, and hence to render an inductive definition of the interpretations of the predicate application pair $\eta, \bar{\eta}$ possible. Of course, the modality is not significant if a statement does not contain predication.[161]

Adding modal comprehension to any given system T still results in a conservative extension of T. As to the specific aim - the search for foundations of Unrestricted Category Theory - the author maintains a cautious attitude, although there are positive applications, showing that the modal extension gains a great freedom (e.g. the extended theory can form the partial classification of all total categories or of all total structures of a given type; a version of the Yoneda's lemma can also be expressed, etc.). It is nevertheless fair to say that the supervenience of the modal structure is not as easy to handle. Additional investigations on modally restricted comprehension are pursued by J. Krajček in [1987a], [1987b].

There are later developments (of 1980 and 1982), where the underlying (classical) logic is extended with an intensional equivalence operator $\equiv$, which is in some sense specifically tailored for the task of living with paradoxes.

The idea is implemented in a joint paper of Aczel and Feferman [1980], where unrestricted comprehension is made consistent by means of an extension of classical logic based on a new intensional operator $\equiv$. According to the authors, $\equiv$ is to be read as: "A is equivalent to B *in consequence of given basic definitions*" and Feferman refers to Behmann's work of 1931 (cf. [Behmann, 1931] and section 3.3) as a source of inspiration.

In order to summarize the axioms, let us assume that we are given a first-order language $\mathcal{L}$. Let $\mathcal{L}^*$ be the expansion of $\mathcal{L}$ with abstraction operator, the new binary operator $\equiv$ with propositional constants $\top$ (true) $\bot$ (false) and the membership $\in$. Let us further define the truth (falsity) operator TA (FA) by $A \equiv \top$ ($A \equiv \bot$); DA (A is determinate) means that A has a definite truth value (true or false).

If S is any theory in the basic language $\mathcal{L}$, S^* is the extension of S in $\mathcal{L}^*$, which, in addition, includes the $\equiv$-logic and naive abstraction in the form

$$\forall x(x \in \{u \mid \varphi\} \equiv \varphi[u := x])$$

(φ arbitrary formula of $\mathcal{L}^*$). The $\equiv$-axioms require that

(i) $\equiv$ is an equivalence relation, preserving the standard logical connectives and quantifiers;[162]

(ii) T and F meet the Kleene weak three-valued logic conditions for standard logical connectives and quantifiers;[163]

[161] In this case $\Box\varphi \leftrightarrow \varphi$ will be provable.

[162] For instance, $(\forall x)(A \equiv B) \rightarrow (\forall x)A \equiv (\forall x)B$.

[163] Logical connectives are interpreted as strict operations, i.e. they are undefined, insofar as at least one arguments is undefined. E.g. one has the schemata

(iii) $T(A \equiv B) \leftrightarrow (A \equiv B)$; $F(A \equiv B) \leftrightarrow DA \wedge DB \wedge (\neg A \equiv B)$;

(iv) $T(A) \leftrightarrow A$ (A atomic); $F(A) \leftrightarrow A$ (A atomic in the ground language $\mathcal{L}$)

It is immediate to see that the paradoxes simply show (in S^*) that there exist fixed points of negation *with respect to intensional equivalence*, i.e. statements satisfying

$$\varphi \equiv \neg\varphi.$$

But this fact does not imply an inconsistency, as it can be seen by means of an elegant model theoretic construction. The crucial step is to interpret $A \equiv B$ as "*A and B have a common reduct*", with respect to a suitable reduction relation $\geq$, which *is generated by associating to the formulas suitable infinitary terms and a corresponding set of rewrite rules.* The consistency proof is then reduced to a Church-Rosser argument, but, since one has infinitary terms and rules around, the confluence proof involves additional difficulties which are controlled with a stratification of $\geq$ into (transfinitely many) layers $\geq_\alpha$.[164] The procedure yields a model of S^* uniformly in any given model of S, and hence one can establish a conservative extension result, granting that S and S^* have the same theorems in the common language.

To sum up the conceptual and methodological aspects, the emphasis of the paper is mainly on the *instrumental role of the principles involved*, showing that the use of certain principles is primarily 'a matter of convenience'. Also, it is stressed that, insofar one is searching for workable version of untyped principles, the logical basis has to be manageable and hence it is better to experiment with *expansions of ordinary logic* than to rely upon some sort of three-valued logic (which is regarded as debilitating).

Applications of type-free principles to constructive foundations

Historically, Feferman [1974a; 1977; 1979] were originally motivated by a specific interest in the logical analysis of Bishop's constructive analysis. Indeed, in [Feferman, 1975] there appear the first formal systems for *explicit mathematics* (EM in short), based on the notions of *partial operation* and *total class(ification)*, where a considerable amount of self-application is allowed. For instance, the intended universe is always a (suitably extended) partial combinatory algebra; so it makes sense to apply an operation f to itself, even if it is not true in general that f converges to a value. On the same par, it makes sense, to consider if a class X is a member of itself, and there exists a universal class V such that $V \in V$.

$$F(A \wedge B) \leftrightarrow D(A) \wedge D(B) \wedge (F(A) \vee F(B));$$
$$F(\neg A) \leftrightarrow T(A) \ ; T(\neg A) \leftrightarrow F(A);$$
$$\ldots$$

[164] According to techniques used by Maas and Girard.

[Feferman, 1975] is in fact the first of a series of papers developing a non-extensional approach to the foundations of constructive or semi-constructive (predicative or borelian) mathematics. As a whole, the project is named *explicit mathematics*, because it insists on the tenet that mathematical objects ought to be explicitly presented by means of computational rules, if one is dealing with functions, or of definitions, if we deal with classes.

A distinctive feature of EM-systems is that, since operations and classes are first-class individuals, they can be classified and operated upon. However, there is a certain asymmetry between classes and operations. Principles ruling operations are primarily computational (even if they can include some abstract functional, like a quantifier), while classes are defined by combining the basic notions by means of logical operations and can be highly non-effective (though many principles are either predicative or produce inductively generated objects of higher and higher complexity).

Formally, in their original intensional version, systems of EM are formulated in a two-sorted language with individual variables x, y, $z \ldots$, classification variables X, Y, $Z \ldots$, predicates symbols for natural numbers N (unary), application Ap (ternary), equality $=$, membership $\in$, individual constants naming combinators and basic number theoretic operations, and in addition constants j, c_n (for every $n \in \omega$) required by the uniform class existence axioms. Prime formulas have the form $t = s$, $t \in X$, $Ap(t, r, s)$(usually shortened as $tr \simeq s$), where terms are variables of either sort or constants. Formulas are inductively generated from prime formulas by closing under logical operations and quantification (of either sort). Due to the freedom of the syntax, one has to carefully choose the set of conditions which are admissible in the class existence principles. *Stratified formulas* are those which contain class variables only in contexts of the form $t \in X$, where t is an individual variable or a constant; *elementary formulas* are those stratified formulas, which contain *no bound class variable*.

An essential aspect of EM is that *the class existence principles are always given uniformly*, by means of an operation of the underlying combinatory algebra. For instance, there is an operation j which takes care of generalized sums, in the following sense: if f is an operation such that fa is a class, for every element $a \in X$, then $j(X, f)$ is a class containing exactly every ordered pair (u, v), where $u \in X$ and $(v \in fu)$.

In EM there are two main comprehension principles under which classes can be consistently required to be closed: elementary comprehension ECA and stratified comprehension SCA. They correspond to special cases of uniform comprehension $\mathcal{F}$-CA

$$(\exists X)(c_A \overrightarrow{Y} \simeq X \wedge (\forall x)(x \in X \leftrightarrow A(Y, x))$$

where A is elementary (or stratified) and c_A is c_k, for some k depending upon the Gödel-number of A.

As it has been readily realized, *this uniformity is essentially responsible for paradoxes*, or, more precisely, for *refutations of classical set theoretic principles*:

- non-extensionality of classes (Gordeev, see [Beeson, 1985]);
- inconsistency (see [Feferman, 1975]) of ECA, join and (a form of) weak power class-existence;[165]
- inconsistency of (some forms of) the axiom of choice [Feferman, 1975; Feferman, 1979];
- inconsistency of the so called strong power-class axiom[166] with ECA [Jäger, 1997].

Similarly, the interplay between operations and classes yields very strong induction principles, whose strength or even consistency is not fully understood.[167] Since the eighties, EM-theories are investigated as potential tools in the area of Logic and Computer Science[168] and as theories of flexible types. The change in perspective has prompted a new formalization of EM, Jäger's *theories of types and names* [Jäger, 1987]. In these systems intensional and extensional aspects coexist: besides individuals, abstract extensional collection or types, are admitted and they are linked by means of a representation relation $\mathcal{R}(x, Y)$, meaning that x represents Y (or is a name of Y). *Every type has a name* and the uniformity of class existence principles amounts to this feature: to each type constructor CS producing a type $CS(X_1, \ldots, X_n)$ uniformly in $X_1, \ldots, X_n$, there corresponds an operation cs such that $cs(x_1, \ldots, x_n)$ is a name of $CS(X_1, \ldots, X_n)$, provided x_i names X_i $(1 \leq i \leq n)$. For inconsistencies in EM as a theory of types and names, see also [Cantini and Minari, 1999].

4.5 Contraction-free and many-valued logics

In the early fifties and sixties, several logicians experiment with type free comprehension, in presence of some form of many valued logics, in particular Łukasiewicz's

[165] The weak power axiom POW has the form:

$$(\forall X)(\exists Y)((\forall x)(x \in Y \rightarrow (\exists Z)(Z \subseteq X \wedge Z = x) \wedge (\forall Z)(Z \subseteq X \rightarrow (\exists U)(U \equiv Z \wedge U \in Y))))$$

[166] The strong power axiom POW^+ has the form:

$$(\forall X)(\exists Y)(\forall x)(x \in Y \leftrightarrow (\exists Z)(Z \subseteq X \wedge Z = x))$$

[167] As an instance, let us only recall the so-called principle UMID of Uniform Monotone Inductive Definition. If

$$Mon(f) := (\forall X)(\forall Y)(X \subseteq Y \rightarrow fX \subseteq fY),$$

UMID claims that there exists an operation fix such that

$$Mon(f) \rightarrow (\exists Y)(fix(f) \simeq Y \wedge fY \simeq Y \wedge (\forall X)(fX \subseteq X \rightarrow X \subseteq Y)).$$

For results on UMID, cf. [Takahashi, 1989], [Rathjen, 1998], [Rathjen, 1999].

[168] Hayashi and Nakano [1988] proposed a constructive logic for computation PX, which was explicitly based on [Feferman, 1979].

many valued logics (see [Skolem, 1957b], [1957c], [1957a],[1960], [1963], [Mow-Shaw-Kwei, 1954], [Chang, 1963], [1965], [Fenstad, 1964], [Brady, 1971]).

Among them, Mow-Shaw-Kwei's work of 1954 studies the contraction rule A_n: from

$$A \to \ldots \to A \to B \quad (n+1 \text{ arrows})$$

infer

$$A \to \ldots \to B \quad (n \text{ arrows})^{169}$$

It is easily shown that Curry's paradox is derivable, as soon as one deals with a system which comprises A_n, modus ponens, abstraction. Also, the author remarks that:

1. the Łukasiewicz systems L_n do have contraction A_{n-1};

2. the Łukasiewicz system $Ł_\aleph$ with infinitely many values does not allow contraction A_n, for each $n \in \omega$.

While it follows that L_3 is inconsistent with abstraction, one can naturally raise *the question whether the theory of sets with unrestricted abstraction in $Ł_\aleph$ can be consistently developed.*

In a philosophical context, Geach [1955] suggested that *a promising route for the semantical paradoxes might be rejecting contraction.* Indeed, aiming at showing that there are semantical *insolubilia* not involving negation, he produces a semi-formal argument, yielding a fixed point of the formula $T(x) \to A$, whence a proof of A, where A is an *arbitrary sentence*, under the assumption that T is a naive notion of truth.[170] In a brief discussion, he notices that besides modus ponens, the paradox exploits "the rule of absorbtion" (that is, infer $A \to B$ from $(A \to (A \to B))$) and that " it might be worth consideration whether this rule should be modified", as "it never occurs in ordinary discourse and we might have a wrong idea of its logical force".

Historically, Geach's short paper marks the appearance[171] in a logical context of the formula involved in Löb's theorem (cfr. [Löb, 1955], [Smorinski, 1985]), modulo the substitution of the naive truth predicate with a standard provability predicate

Mow-Shaw-Kwei's problem is tackled in 1957 by Skolem [1957b], who presents a partial solution. After remarking that unrestricted comprehension is inconsistent with each finite-valued Ł-logic, he goes on to show that, if comprehension is restricted to *quantifier-free* conditions, then naive comprehension is consistent modulo $Ł_\aleph$. Skolem's approach is conceptually important, since *he observes that the problem of type-free comprehension can be recast as a problem in Analysis.* In Skolem's words, the consistency issue amounts to the

[169] Brackets are restored by associating to the right.
[170] Thus satisfying $T(\lceil A \rceil) \leftrightarrow A$.
[171] Is this the first time?

> "the investigation of a mathematical function $\mathcal{E}(x, y)$, which assumes as values rational numbers ≥ 0 and ≤ 1, where x, y must range over individuals of a domain. Therefore the comprehension axiom can be stated within the usual predicate calculus in the form
>
> (14) $(\exists y)(\forall x)(\mathcal{E}(x, y) = U(x))$
>
> where now $U(x)$ is a mathematical function, which is built up by means of minimum and maximum operations".

Indeed, Skolem remarks that, if $A(p)$ is a sentential function built up by means of the logical connectives $\wedge$, $\vee$, $\neg$, $\rightarrow$, which are interpreted à la Łukasiewicz as the functions $min(x, y)$, $max(x, y)$, $1 - x$, $min(1, 1 - x + y)$ (in the given order), then $A(p)$ is piecewise-linear and defines a polygonal curve, whose pieces have rational derivatives and whose edge-points are rational (and the geometrical representation is easily generalized in case of formulas with several variables). Since the function $p \mapsto A(p)$ maps the real interval $[0, 1]$ into itself and is continuous, there exists at least a fixed point h solving the equivalence

(15) $h \leftrightarrow A(h)$

Now Skolem is able to prove that (14) can be solved for $\mathcal{E}$. First, all propositionally definable functions of $Ł_\aleph$ with finitely many variables in the list x, x_1, x_2, ... are enumerated as U_1, U_2, ... Then, by crucially exploiting (15), an infinite sequence $x_1, x_2, \ldots$ of individuals beginning with 0 is inductively defined, which meets the following requirements:

(i) if $x_1, \ldots, x_m$ are already introduced, $U_n(x, x_1, \ldots, x_m)$ is a propositional function in the fixed enumeration, there is a new individual $f_n(x_1, \ldots, x_m) = x_{m+1}$ and

(ii) it is possible to fix stepwise the values of all atoms of the form $x \in y$, so that the value of $x \in y$, for $y = f_n(x_1, \ldots, x_m)$, coincides with the value of $U_n(x, x_1, \ldots, x_m)$.

The paper also contains a discussion of the difficulties arising with quantifiers and of possible significant applications to the foundations of mathematics (this is apparently Skolem's main motivation for the whole work[172]).

Skolem's partial result is improved by [Chang, 1963]. Using the Brouwer's fixed point theorem and a version of the compacteness theorem for the logic $Ł_\aleph$, Chang is able to show the consistency of the set of all sentences of the form

(16) $\exists y \forall x(x \in y \leftrightarrow \varphi(x, y))$

where $\varphi(x, y)$ is an arbitrary first order formula with at most the variables x, y free, whose only predicates are $\in$, $=$. Consistency is understood semantically, as

[172] Similar issues are also discussed in the partially overlapping [Skolem, 1957a].

the existence of a model $M = \langle S, e\rangle$ assigning value 1 to each instance of (16); here S is a non-empty set, e is a function from $S \times S \to [0,1]$ (= is understood as true identity on S) and formulas are evaluated according to the $Ł_{\aleph}$-interpretation of logical operations. Further results of [Chang, 1963] and [Fenstad, 1964] concern consistency when parameters are present, though they are allowed to occur in restricted contexts.

Skolem's ideas will be further developed in the frame of fuzzy arithmetic more than forty years later (see [Hájek *et al.*, 2000]).

In the early sixties Skolem obtains another consistency result, which shows that $Ł_{\aleph}$ is not necessary for consistency, as soon as implication is not involved in the comprehension schema: $Ł_3$-logic, extensionality and type-free comprehension can consistently live together, provided the comprehension formula contains only negation and conjunction [Skolem, 1960], [Skolem, 1963]. [Brady, 1971] slightly improves upon this result: consistency holds even if the universal quantifier is allowed in the comprehension formulas. Roughly, the point is that, if you do not have implication, the underlying evaluation schema is *monotone*; so Gilmore's inductive method for partial set theory applies (see above and Feferman [1984]).

Consistency of full naive comprehension *with parameters and abstraction operator* in the infinite valued predicate logic is eventually proved in the 1979 paper of White, who outlines an elaborated proof combining proof theoretic techniques (partial normal form theorem for a finite variant of the complete infinitary axiomatization of predicative Ł–logic) with a final Henkin type argument.[173] He further notices that inconsistency arises with extensionality.

Grišin [1974], [1981] develops a "non-standard logic" without contraction and proves that the unrestricted type free comprehension principle is consistent by using a constructive cut-elimination argument.[174] Interestingly, the same author notices in [Grišin, 1981] that *extensionality (in the usual formulation) ruins consistency, since it implies contraction, in presence of full comprehension.*

As to the possible applications, one should observe that naive comprehension does not attain enough deductive power, in order to recover the role of a 'grand logic' (in the old-fashioned foundational sense), though there are some attempts toward sensible blends or non-trivial refinements, with the help of the now flourishing tree of substructural logics. Systems with naive comprehension and some form of contraction free logic, being very expressive, may have some interest *not for the foundations of mathematics, but for computability theory.*[175]

[173] The argument is involved and we have not been able to check the proof in all details.

[174] [Grišin, 1981] defines an algebraic semantics for the given logical calculus and proves a corresponding completeness theorem; this is of interest for the history of substructural logics, since Grišin's calculus is a version of linear logic with weakening, i.e. affine logic without exponentials (see [Girard, 1987a]) As far as we are aware of, at present, there exist only proof-theoretic consistency proofs for Grišin's system (except for Komori's illative version in [Komori, 1989], which is any way weaker, being based upon intuitionistic affine logic).

[175] In recent years, Girard [1998] has proposed a logical characterization of polynomial time computability by means of naive comprehension together and a suitably modified linear logic (the so called 'light linear logic').

5 THE PROBLEM OF SEMANTICAL CLOSURE: PHILOSOPHICAL AND TECHNICAL VARIATIONS

Though there is a variety of motivations, scattered approaches and sketchy solutions, a strong impulse derives from the philosophy of language and from attempting *a rational reconstruction of natural language semantics*, against the Tarskian negative verdict.

Examples of philosophical diagnoses of semantical paradoxes intertwined with logical analysis are to be found in van Fraassen's papers [1966]–[1972], in R. L. Martin's collective book [Martin, 1970] and in C. Parsons's essay [1974] on the Liar paradox. Some papers are indeed preparatory of significant technical work,[176] which is carried out only a few years later.

Clearly, the shift of emphasis from formalized languages to natural ones explains why the resolution strategies tend to reject the hierarchy of truth predicates, aiming instead at formal frames, capable of the highest degree of semantical closure. Furthermore, in order to avoid inconsistencies, the new investigations generally set up *some form of three valued semantics (gap solution)* against bivalence. The idea is that one should add a third value u ('undefined' or 'undetermined'), so that negation $\neg$ could consistently satisfy $(\neg u) = u$ (this equation being implicitly regarded as characteristic of paradoxical claims), and thus avoid contradictions.

The fixed point condition on negation is only a minimal requirement and it becomes important to choose semantical frameworks, which allow partially defined predicates and yet stay close to ordinary classical reasoning. A straightforward move is to replace classical truth tables – regarded as truth functions $F : \mathbf{2}^m \to \mathbf{2}$ of the appropriate arity – with their three-valued extensions in the sense of [Kleene, 1952], which satisfy the conditions:

(i) the set $\mathbf{2} = \{t, f\}$ of classical truth values is replaced by $\mathbf{3} := \{t, f, u\}$, u stands for the third value 'undefined' (or 'undetermined') and $\mathbf{3}$ is partially ordered by letting $u \leq_T t$ and $u \leq_T f$;

(ii) they are $\leq_T$*–monotone.*[177]

The simplest three-valued truth schema is given by *the weak Bochvar-Kleene tables* (see [Bochvar, 1937]), the unique extensions of the two valued ones obtained by supplying u as soon as any of the arguments is u. Alternatively, *the strong Kleene tables* are the strongest possible monotone extensions of the classical truth functions.[178]

[176] Let us only mention the papers of Skyrms, Herzberger, Martin in [Martin, 1970].

[177] A three-valued m-ary function $F : \mathbf{3}^m \to \mathbf{3}$ is monotone if, whenever $x_i \leq_T y_i$ where $1 \leq i \leq m$, $x_i, y_i \in \mathbf{3}$, then $F(x_1, \ldots, x_m) \leq_T F(y_1, \ldots, y_m)$.

[178] They can explicitly described by the following stipulations: a conjunction, disjunction, or implication, one of whose components is assigned u, receives t or f according to the case that the truth value is already classically fixed, independently of whether u is replaced by t or f, and is given value u otherwise.

Both semantical schemata allow extensions with three-valued quantifiers. Let $\forall^c$, $\exists^c$ be the standard classical quantifiers.[179] In the case of weak Bochvar-Kleene logic, if $\Phi : D \to \mathbf{3}$, and Q^c is $\forall^c$ or $\exists^c$, let

$$Q^{wk}\Phi = \begin{cases} u, & \text{if } \Phi(a) = u, \text{ for some } a \in D; \\ Q^c\Phi, & \text{else.} \end{cases}$$

As to strong Kleene logic, the quantifiers $\forall^{sk}$, $\exists^{sk}$ satisfy:

$$\forall^{sk}\Phi = \begin{cases} t, & \text{if } \Phi(a) = t, \text{ for every } a \in D; \\ f, & \Phi(a) = f, \text{ for some } a \in D; \\ u, & \text{else;} \end{cases}$$

$$\exists^{sk}\Phi = \begin{cases} t, & \text{if } \Phi(a) = t, \text{ for some } a \in D; \\ f, & \Phi(a) = f, \text{ for all } a \in D; \\ u, & \text{else.} \end{cases}$$

Kleene's three-valued logics, though quite natural, cannot justify the observation[180] that, even if paradoxical sentences must be regarded as semantically defective, *the lack of truth value is the result of arguments*. So one ought to define systems where classical logical validity and logical consequence still make sense, and yet one can deal with possibly semantically undefined sentences. For instance, one would like to maintain the tautological character of $L \vee \neg L$, even if L is neither true nor false; but this is plainly impossible with Kleene's three-valued logics.

In this respect, a solution is given by the supervaluation semantics of [van Fraassen, 1966]– [van Fraassen, 1972].[181] Roughly, the basic idea is this: a sentence which contains *a partially defined predicate* is made true if it is true *under all possible classical completions* (compatible with the available definite information).

More precisely, assume that $\mathcal{L}$ is an elementary language where for each sentence φ there exists a name (i.e. a closed term) $g(\varphi)$ designating φ and there exists a predicate T, to be interpreted as a truth predicate for the entire $\mathcal{L}$.

Let V be a three-valued map from the set of all sentences to the set of truth values $\{t, f, u\}$. Now the *supervaluation*[182] *induced by the three-valued valuation* V is defined as the unique three-valued valuation V^{sv} such that

$$V^{sv}(\varphi) = \begin{cases} t\,(f), & \text{if } V'(\varphi) = t\,(f), \text{ for all classical } V' \text{ extending } V; \\ u, & \text{otherwise} \end{cases}$$

Clearly $V^{sv}(\varphi \vee \neg\varphi) = t$, even if no valuation V is defined on φ (in the sense that it assumes a boolean value). It then follows that *the new semantics is not*

[179] Regarded as functionals acting on the space $D \to \mathbf{2}$; if $\Phi \in D \to \mathbf{2}$, $\forall^c\Phi = t$ if and only if Φ is everywhere t, and $\exists^c\Phi = t$ if and only if Φ is somewhere t.

[180] Cf.[van Fraassen, 1966],[van Fraassen, 1968].

[181] Originally suggested as a tool for providing a logical analysis of presupposition and a systematic interpretation of non-denoting singular terms.

[182] The definition given here admits possible variants; for instance one could quantify on all partial valuations, and not only on classical ones.

compositional (or truth-functional): e.g. the truth value of a disjunction cannot be computed from the truth values of its disjuncts. On the other hand, supervaluations allow truth-value gaps (e.g the Liar) and yet are conservative with respect to classical reasonings (in the sense that supervaluations do not affect the truth value of sentences not involving truth, but, if all classical valuations assign value t to an arbitrary sentence φ, so do all supervaluations).

A final preliminary question is that, if we accept semantical gaps, Tarski's schema $\varphi \leftrightarrow T(g(\varphi))$ cannot be stated unrestrictedly for arbitrary φ. In other words, it is not even clear how to state a sensible adequacy condition for truth, in accord with a partial semantics. A reasonable answer, which is actually implemented in Kripke, Martin-Woodruff, Kindt, is in [van Fraassen, 1970], [van Fraassen, 1971], p. 167.[183] The basic observation is that the existence of truth value gaps still retains an important consequence of the schema, i.e. that, for arbitrary sentence φ, φ and $T(g(\varphi))$ have the same value and hence are *inferentially equivalent*. Hence T should be regarded as an adequate truth predicate if and only if the following inferences are sound for T, given any arbitrary sentence φ:

- (T-intro), from φ, infer $T(g(\varphi))$;
- (T-elim, from $T(g(\varphi))$, infer φ.

The problem of semantical closure then becomes: find a formal elementary language $\mathcal{L}$[184] with a truth predicate T, a naming injective operation g mapping sentences of $\mathcal{L}$ into closed terms of $\mathcal{L}$, and a valuation W such that, for every sentence φ,

(17) $W(T(g(\varphi))) = W(\varphi)$

Clearly, if such a language is proven to exist, there is a language representing its own truth in (an inferentially) adequate fashion.

5.1 Two solutions based on monotone semantics

The problem is not solved until 1975, when we can find three different solutions, due to Kindt [1978b], Kripke [1975], and Martin and Woodruff [1975].[185]

We summarize the common background and features of these approaches (without literally following them); then we'll consider some details pertaining to each contribution. The starting point is an elementary language $\mathcal{L}$, which comprises

[183] But a general form of it is also in Fitch [1964], p. 499.

[184] This restriction is not trivial: for combinatory languages the solution was already there, as we know from [Fitch, 1964].

[185] These works are apparently independent. Kindt's paper appears in 1978; but the author states in [Kindt, 1978a] that the basic ideas were presented at a conference in Bielefeld already in December 1974. Kripke's paper was read in December 1975 at a meeting of the American Philosophical Association. As a matter of fact, the mathematical tools are already applied in Fitch, see 4.2, [Gilmore, 1974](presented in 1967 at UCLA), [Myhill, 1975], [Nepeivoda, 1973], Chauvin's thesis of 1974, [Feferman, 1974a] (presented at the Kiel Symposium of August 1974), [Scott, 1975](which refers to Fitch, Feferman) .

a special monadic predicate symbol T for truth, besides additional relation and function symbols. The T-free part of the language is endowed with an interpretation $\langle D, I\rangle$, where D is a non-empty set and I is an interpretation function in the standard sense; so it assigns elements of D to individual constants, n-ary relations on D to n-ary predicate symbols and n-ary functions of $D^n \to D$ to n-ary function symbols. Further, if r is a closed term of $\mathcal{L}$, there exists a unique $I(r) \in D$, which is the value of r in D. Henceforth $\langle D, I\rangle$ remains fixed. It is also convenient to assume that $\mathcal{L}$ comprises distinct names for each element of D[186] and that the full syntax of $\mathcal{L}$ may be represented in $\mathcal{L}$. As usual, if φ is a sentence of the given language (possibly containing T), $\ulcorner\varphi\urcorner$ will stand both for the closed term designating φ and the corresponding object of D.

Let $\mathbf{3} := \{t, f, u\}$ be the set of truth values, together with the partial ordering $\leq_T$ of the previous subsection. $\leq_T$ naturally induces a *pointwise partial ordering* on the space $\mathbf{3}^{D^n}$ of all n-ary functions from D^n to $\mathbf{3}$ ($n \geq 1$): if F, G are elements of $\mathbf{3}^{D^n}$, let

$$F \leq_T G \Leftrightarrow (\forall \vec{a} \in D^n)(F(\vec{a}) \leq_T G(\vec{a}))$$

In particular, $\leq_T$ is inherited by the set of *partial subsets* of D, which can simply be identified with functions of the space $D \to \mathbf{3}$, or, equivalently, with pairs $X = \langle X^+, X^-\rangle$ of disjoint non-empty subsets of D,[187] where X^+ (X^-) is named *the extension (antiextension)* of the set X. Partial sets are the natural candidates for interpreting the truth predicate T.

Given these preliminaries, the solutions presented by [Kindt, 1978b], [Kripke, 1975] and [Martin and Woodruff, 1975] essentially amount to the choice of a specific semantical schema $\mathcal{S}$, which inductively assigns a truth value $V^{\mathcal{S}}_X(\varphi)$ to arbitrary sentences (once one has chosen a partial subset X for T and stipulated that, for arbitrary closed term r, $V^{\mathcal{S}}_X(T(r)) = X(I(r))$).

Martin and Woodruff[1975][188] stick to the Bochvar-Kleene schema (henceforth in short $\mathcal{S} = WK$) and argue non constructively. First of all, they notice that

(18) $X \leq_T Y \Rightarrow V^{wk}_X \leq_T V^{wk}_Y$

Then they consider the class PR of all WK-valuations "partially representing the truth", i.e. those satisfying the closure condition

(19) $V^{wk}_X(T(\ulcorner\varphi\urcorner)) \leq_T V^{wk}_X(\varphi)$.

They further observe that every $\leq_T$–chain of PR-valuations has an upper bound in PR, whence by Zorn's lemma *there exists a maximal valuation* $V^\infty \in PR$. By

[186] So, strictly speaking, we should make the dependence on D explicit and write $\mathcal{L}_D$ instead of $\mathcal{L}$.

[187] In this presentation, we could define the ordering on partial sets by:

$$X = \langle X^+, X^-\rangle \leq Y = \langle Y^+, Y^-\rangle \Leftrightarrow X^+ \subseteq Y^+ \text{ and } X^- \subseteq Y^-.$$

[188] We disregard the fact that their proposal deals with a many-sorted language and we omit the philosophical context, the so-called category solution, for which we send the reader to [Martin, 1970].

maximality, (18) and (19), V^∞ also satisfies $V^\infty(\varphi) \leq_T V^\infty(T(\ulcorner\varphi\urcorner))$ and hence V^∞ yields a solution to (17).

Kindt [1978b] is motivated by investigations in the field of theoretical linguistics and aims at a canonical general method for introducing truth predicates. His starting point is the natural intuition that the truth value of $T(\ulcorner\varphi\urcorner)$, must be *reducible* to the truth value of φ, and that, in general, every decision about the denotation of T must be grounded upon statements holding for the base structure $S = \langle D, I\rangle$ ([Kindt, 1978b], p. 361). The intuition is made precise by defining a relation of *reducibility*[189] and a corresponding notion of *groundedness.*[190] sentence Roughly, a sentence φ, possibly involving the truth predicate, is grounded if φ is the root of a well-founded[191] reducibility tree, whose leaves are labelled by true formulas in the T-free language. Of course, it is not difficult to show that the set $G(S)$ of grounded sentences (with respect to S) can be inductively defined (apply the clauses).

The notion of groundedness is then crucially used by Kindt, together with the strong Kleene three valued logic, for characterizing a truth interpretation W, which is recursively generated by closing under the rule

$$V^{sk}_W(\varphi) = t\,(f) \Rightarrow W(T(\ulcorner\varphi\urcorner)) = t\,(f),$$

and hence uniquely determined as *minimal* among the set of all truth extension satisfying (17).

5.2 Kripke's "Outline of a theory of truth"

Kripke [1975] offers a unifying conceptual frame for reconsidering and evaluating previous scattered contributions arising in philosophical contexts.[192] Technically, his proposal is a genuine extension of Tarskian semantics, it provides a mathematical classification of self-referential statements and it also subsumes [Kindt, 1978b] and [Martin and Woodruff, 1975] as special cases.

A few words on Kripke's philosophical starting points. First of all, he argues that a sentence φ is always *meaningful*, but the act of using it to express a proposition is potentially risky, since in unfavorable circumstances[193] it can fail to have a

[189]Similar notions are to be found in connection with the construction of the canonical search tree used for the completeness proof à la Schütte and in the related semantical concept of Schütte valuation, cf. [Schütte, 1960], [Schütte, 1977], [Girard, 1987b].

[190]Of course, a similar notion is used by [Kripke, 1975], p. 695, which refers also to [Herzberger, 1970] and to previous unpublished work of Herzberger and Katz.

[191]Kindt calls these well-founded trees 'finite'; but this is rather odd, these trees being infinitary in general.

[192]He actually complains that previous contributions in the field give only mere suggestions and not precise concepts. According to him, the (then yet) unpublished paper [Martin and Woodruff, 1975] is the only exception, recognizing it as a (partial) solution, in the sense that maximal fixed points are shown to exist. He also refers to work of Fitch, Gilmore, Feferman as fairly close to his own approach, though related to set-theoretical paradoxes.

[193]Empirical, and not only derived from self-referential tricks, as in the paradoxes. A related observation has been made earlier; the role of accidental self-reference is noticed in [van Fraassen, 1970].

definite truth value. Consequently, we must live with gaps and partially defined predicates[194] and there is no hope for a syntactical regimentation of the language; a solution to paradoxes must necessarily follow a semantical route.

In the second place, Kripke addresses the question: how could we explain the word 'true' to someone who does not yet understand it? The answer is similar to the one presented by [Kindt, 1978b]. But Kripke explicitly describes a model of this attempt to ground sentences with truth predicate occurrences upon sentences that do not speak about truth, as an *evaluation process developing in stages* and approximating the extension (and the antiextension) of the (necessary partial) truth predicate. The approximation process is naturally cumulative or monotone: no proper truth value t or f, once established, is discarded, and a statement can only pass from the undefined state to a definite truth value. First of all, Kripke deals with a language $\mathcal{L}$ for an acceptable structure (say, of domain D) in the sense of Moschovakis [1974]; this ensures that $\mathcal{L}$ has enough encoding abilities for the formalization of syntax. Then he proceeds to define a jump operation $J^{\mathcal{S}} : \mathbf{3}^D \to \mathbf{3}^D$ such that, if $Sent_D$ is the set of all codes of $\mathcal{L}$–sentences and $a \in D$,

$$J^{\mathcal{S}}(X)(a) = \begin{cases} t, & \text{if } a = \lceil\varphi\rceil, \text{ for some sentence } \varphi \text{ with } V^{\mathcal{S}}_X(\varphi) = t; \\ f, & \text{if } a = \lceil\varphi\rceil, \text{ for some } \varphi \text{ with } V^{\mathcal{S}}_X(\varphi) = f, \text{ or } a \notin Sent_D; \\ u, & \text{else.} \end{cases}$$

$J^{\mathcal{S}}(X)$ yields the sentences whose truth value can be grounded assuming that X is the denotation of truth. Kripke is the first to notice explicitly that, if $\mathcal{S}$ is any $\leq_T$–monotone scheme for handling truth-value gaps, the induced jump operator is monotone, and hence there are plenty of fixed points, i.e. partial sets M such that $J^{\mathcal{S}}(M) = M$;[195] clearly they solve (17) and hence there are many possible candidates for semantically closed predicates.

Of course, a privileged role is still played by the *minimal fixed point* $\mathcal{O}_{\mathcal{D}}$. Mathematically, $\mathcal{O}_{\mathcal{D}}$ is generated by iterating the jump operator over the ordinals, starting from the empty partial set, and it directly implements the informal idea of grounding and approximating truth from below. By a standard set theoretic argument, the process becomes saturated at a certain point (*the closure ordinal*), and hence it reaches the least set closed under $J^{\mathcal{S}}$, which is in fact the least fixed

[194] As before (with a little violence against the text of [Kripke, 1975] and for expository reasons), we regard partial logic and three-valued logic as interchangeable, thus relying upon the space $\mathbf{3}$, ordered by $\leq_T$. But Kripke firmly claims that the third value u only stands for lack of a proper value; so it should not be regarded as a departure from classical logic.

[195] For the existence, one might appeal to some generalized form of Tarski's fixed point theorem; for instance, see [Markowski, 1976], where the fixed point theorem is proven for monotone maps of a chain-complete poset into itself or for maps of a strictly inductive poset into itself, satisfying $f(x) \geq x$. An abstract setting explicitly devised for Kripke's paper is to be found in section 2 of [Visser, 1989] and is influenced by work done in denotational semantics. Visser defines the notion of *coherent complete partial ordering (ccpo)*. A ccpo D is a poset, where every consistent subset has a sup; a subset X of D is consistent, provided any two elements of D have an upper bound in D. If D is a ccpo, monotone maps from D into D have fixed points, and the set of all fixed points of a given monotone map yields a ccpo. Related work is in [Fitting, 1986].

point. So it is natural to call all sentences with a definite truth value in $\mathcal{O}_{\mathcal{D}}$ *grounded*.

The plurality of fixed points is nicely exploited to define a semantical classification of ungrounded sentences. Consider the informal statements

(‡) (‡) is true.

(♮) (♮) is not true;

Clearly both sentences are ungrounded; there is however an intuitive difference between (‡) and (♮). We cannot apparently derive a contradiction by assuming that (‡) is true (or false); and in fact there are fixed points which make (‡) true and others where (‡) is false. But exactly this is forbidden by the second statement, which is in fact a version of the Liar and is undefined in every fixed point. So we can define a sentence as *paradoxical* if it is *undefined in every fixed point*.

These remarks open up the place to a closer study of fixed points and a consequent subtler classification of ungrounded non-paradoxical sentences.[196] In this respect, Kripke mentions two noteworthy classes of fixed points. As we know already from [Martin and Woodruff, 1975], there exist by Zorn's lemma plenty of maximal fixed points.[197] They deserve attention, since they converge on as many sentences as possible. On the other hand, there exist the so-called *intrinsic fixed points*: they assign a value to a sentence φ, only if this value remains the same in every other fixed point which is defined on φ. Interestingly, *intrinsic fixed points do form a complete lattice* and hence there exists the largest intrinsic fixed point $\mathcal{I}_D$. $\mathcal{I}_D$ can be regarded as an improved notion of truth, which is optimal,[198] once we extend the minimal fixed point with ungrounded sentences, but under the constraint that *no arbitrary choice* is made in assigning truth values. Quite naturally, we say that an arbitrary sentence φ has an intrinsic value (or simply is intrinsic) if φ has a value in $\mathcal{I}_D$. This means that φ is defined and has the same value in every maximal fixed point. For instance, if we let

(♯) Either (♯) is true or its negation is true,

then (♯) is ungrounded and intrinsically true. On the other hand, if we let

(♭) either (‡) is true or its negation is true,

one verifies that (♭) is ungrounded, it is true in every fixed point where (‡) has a truth-value, but it cannot be intrinsically true (as there are maximal fixed points where (‡) is undefined).

[196] Warning: this type of results will not be invariant under the choice of a monotone three valued schema $\mathcal{S}$. As Kripke himself remarks, there may occur sensible differences, once one varies $\mathcal{S}$ choosing a special class of supervaluations ([Kripke, 1975], p. 711) or the weak Kleene Schema (cf. [Cain and Damnjanovic, 1991]).

[197] E.g., if the language is countable, there are uncountably many maximal fixed points; for a proof, see [Cantini, 1989].

[198] In fact, this is connected with the notion of optimal fixed point coming from computer science. It is also possible to show that the cardinality of intrinsic fixed points is the highest possible with respect to the cardinality of the domain of the interpretation; cf. [Cantini, 1989].

5.3 Kripke's offsprings

Last but not least, Kripke's paper suggests applications of his own theory of truth to recursion theory.[199] For instance, the Π^1_1-sets of numbers coincide with the extensions of monadic formulas, and Δ^1_1-sets are the extensions of totally defined monadic formulas. So the language of self-referential truth gives a notation-free version of hyperarithmetical sets. More generally, Kripke naturally conjectures a generalization of the result to the case of languages with partial truth, based on acceptable structures.[200] Not only Kripke's construction is an application of the theory of inductive definition, but it will later result that Kripke's fixed point approach yields "a natural formulation of many elementary facts about inductive definition" and hence one can naturally speak of *a two-sided connection between the two fields* .[201]

Concerning the philosophical import of his paper, Kripke is rather cautious: certainly, his model improves upon the hierarchical solution of the Liar paradox, since it satisfies more closure properties than Tarski's.[202] However, there are *definite limits to what we can express in the Kripkean language.* Typically, it is shown that the Liar is not true in any fixed point, but *this very fact cannot be expressed in the language*: so one cannot refrain from saying, with Kripke's words, that "the ghost of the Tarski hierarchy is still with us (p. 714)".

A fortiori, it should also be considered that typical semantical notions in Kripke's work (e.g. 'intrinsic', 'paradoxical', etc.) have *a high complexity from the recursion-theoretic point of view* and go beyond the borders of Kripke's language.

The main results in this direction are due to Burgess [1986], [1988]. If $S := K(VF)$ is a label designating the Kleene strong three-valued (supervaluation) schema (respectively), let O^+_S be the least S-fixed point; further, let V^+_S (V^-_S) be the set of (Gödel-numbers of) sentences which are true (false) in some S-fixed point (over the standard model of PA). Then it can be shown:[203]

(a) V^+_K, V^-_K and $V^+_K \cap V^-_K$ are Σ^1_1-complete;

(b) $\omega - (V^+_K \cup V^-_K)$ is Π^1_1-complete;

[199] Henceforth we assume that the reader is acquainted with the basic notions touching the first levels of the analytical hierarchy, as can be found in [Rogers, 1967], ch.15 or in [Odifreddi, 1989], ch. IV.2.

[200] A proof of this claim is credited to L. Harrington (footnote, p. 716). There is an implicit technical assumption behind these results, i.e. the given language must have finitely many non-logical symbols. The reader should not forget earlier work – in foundational vein – due to [Lorenzen and Myhill, 1959] [Nepeivoda, 1973]; related results in the context of systems based on type-free comprehension were announced in Feferman 1975 [1974a] and have been later established in published form by several people, e.g. [Cantini, 1980], [Flagg and Myhill, 1987b], [Visser, 1989].

[201] As far as we know, this has been explicitly underlined by Visser [1989] and systematically pursued in [Cantini, 1989], [1996], where it has been shown that there is a virtuous link relating Kripke's construction, Frege structures and admissible set theory.

[202] Perhaps, it is appropriate to say that Kripke's theory is to Tarski's as ZFC is to the theory of types.

[203] Some of these statements are credited by Burgess to Kripke himself.

(c) $V_K^+ - V_K^-$ is complete among the sets which are differences of Σ_1^1-sets.

(d) The set I_K^+ of (Gödel-numbers of) sentences true in all intrinsic K-fixed points is complete among the sets which are Σ_1^1 in-a-Π_1^1-parameter (actually in Kleene's $\mathcal{O}$).

By way of comparison, even if O_K^+ and O_{VF}^+ have same complexity (i.e. are Π_1^1-complete), Burgess has shown:

(a) V_{VF}^+ is Σ_2^1-complete;

(b) I_{VF}^+ is at least Π_2^1 and at most Σ_2^1-in-a-Π_2^1-parameter.

In conclusion, it is difficult to overestimate the influence of Kripke's paper on subsequent work. The 1975 paper is followed by plenty of variations and generalizations, that cannot be referred to here for lack of space (e.g. [Yablo, 1982], [Yablo, 1985], [McGee, 1991], [Pelham and Urquhart, 1994], with a resolution of the contradiction in Russell's original substitutional theory); let me also mention the natural extension of Kripke's work by means of the Dunn-Belnap 4-valued logic by [Visser, 1984] and [Woodruff, 1984] (see the remark below).

Apart from its philosophical import, [Kripke, 1975] leaves open *interesting mathematical questions*, which range from axiomatic and proof-theoretic investigations to recursion theoretic.[204]

Kripke's work also marks a fresh start in formal semantics since Tarski, and this is witnessed by the number of monographs in the subject that will appear in the eighties and in the nineties (and still appear).[205]

Generalized Kripke fixed points

Let us briefly summarize the main content of [Visser, 1984], [Woodruff, 1984]. Apart from inessential differences, the construction amounts to replacing the space of truth values in the definition of Kripke's jump with the lattice $\mathbf{4}$ of the subsets of $\{t, f\}$. Then a candidate for truth is simply a map of $\mathbf{4}^D$, or more simply a pair (X^+, X^-) of subsets of D. If the pair is disjoint (covers D), it is consistent (complete).

Then the Kripke jump is still monotone from $\mathbf{4}^D$ into itself, and the structure of fixed points becomes a complete lattice, endowed with *the natural involution*, (essentially) sending X^+ (X^-) into the complement (relative to D) of X^- (X^+).

[204] See [Feferman, 1991], [Reinhardt, 1985], [Reinhardt, 1986], [Cantini, 1989], [Turner, 1987] and others. There is also an application of Kripkean ideas to the discussion of the notion of *proper class* in Maddy's paper [1983].

[205] Here is a tentative list: [Barwise and Etchemendy, 1987], [McGee, 1991], [Turner, 1990], [Gupta and Belnap, 1993], [Koons, 1992], [Simmons, 1993], [Yaqūb, 1993], [Halbach, 1996], [Cantini, 1996], [Chapuis, 1996].

In this structure the least fixed point has a dual, the greatest fixed point MAX_D, which is complete but inconsistent (as it assigns to the Liar both t and f).[206]

The structure of the lattice of generalized truth predicates has been further investigated in [Fitting, 1986], [Fitting, 1989]. If D is a model of PA or a combinatory algebra, or else associated with infinitary logics of truth, it has been proved that the lattice is non-modular ([Minari, 1987], [Cantini, 1989]).

Appendix on semantical closure: minimalism and logical difficulties

One might tackle the problem of semantical paradoxes, holding that the naive theory of truth where one assumes the T-schema $T(\lceil\varphi\rceil) \leftrightarrow \varphi$, for arbitrary φ, is essentially tenable and sufficient. One has only to filter out those instances, which are responsible for the contradictions. This view (systematically held for instance by the philosopher P. Horwich) presupposes that the task implicit in the minimalist standpoint is rationally feasible. However, here is a difficulty pointed out by V. McGee [1992]: *there is no effective method to decide whether a specific instance of the T-schema is consistent* (choose a fixed point ψ of $T(x) \leftrightarrow \varphi$, where φ is arbitrary and observe that φ is equivalent to $T(\lceil\psi\rceil) \leftrightarrow \psi$).

Interestingly, a similar difficulty affects the attempt of amending the naive comprehension schema as well as the schema for introducing contextual definitions underlying the Fregean definition of natural number (this remark is due to Boolos and Heck; see [Boolos, 1998, pp. 232-233]).

5.4 Self-referential truth and type free systems: a unified framework

In the previous sections 4.2-4.5, 5.1-5.3 we have summarized *in parallel* some relevant work about type-free principles and formal semantics. Their common conceptual and technical roots are made explicit in [Feferman, 1984], which outlines a general setting for treating self-referential constructions both at the level of naive semantics and type-free comprehension. Also, it yields an analysis of the features that make the Liar and the Russell paradox possible. These can be can grouped together into three categories:

(i) *syntax:* there are *naming* facilities in the language (e.g. individual terms $\lceil\varphi\rceil$ representing sentences φ, as well as individual terms $\{x \mid \varphi\}$ representing classes defined by arbitrary φ);

(ii) *logic*: ordinary axioms and rules are accepted;

(iii) *basic principles*: these typically amount to the naive comprehension schema and the unrestricted truth schema $T(\lceil\varphi\rceil) \leftrightarrow \varphi$.

[206] The dual Kripkean approach is a way of making sense of the view that paradoxes force us to accept inconsistencies. For different approaches leading to the investigation of the so-called 'paraconsistent logics', see [Priest, 1989].

Strictly speaking, (i)-(iii) can be consistently met and the Liar does not necessarily appear yet: Gupta [1982, p. 14] has shown that *there are classically interpreted languages containing quotation names for their sentences and their own truth predicate*, which satisfy unrestricted T-schema for sentences. Hence satisfaction or the use of a truth predicate with substitution or concatenation functions, (e.g. in the style of Quine and [Smullyan, 1957]) are necessary.

Given this warning, a solution route to paradoxes can reasonably attempt to find a consistent theory or a model with *restrictions* on naming, logic or basic principles. But, as already stressed in [Aczel and Feferman, 1980], the strategy followed by Feferman is to neutralize paradoxes *by means of extensions of first order logic*, in a way that compensates the critical adoption of unrestricted self-referential principles.

The starting point is a language $\mathcal{L}_0$, which only involves *total* relations, with a corresponding structure $\mathcal{M}_0$ of the same similarity type. It is further assumed that the language has means to encode a first order syntax and that all structures satisfy a minimal theory S_0 which grants it. So there will be terms encoding finite sequences (for which we use $(x_1, \ldots, x_k)$) and Gödel-numberings $\ulcorner\varphi\urcorner$ of formulas φ.

The main problem is now to investigate the semantics for expansions $\mathcal{L}^+$of $\mathcal{L}_0$ with $k+1$-ary predicates T_k (where $k \in \omega$). Roughly, the atomic formula $T_k(x_1, \ldots, x_k, z)$ is read as: " $(x_1, \ldots, x_k)$ *satisfies* z". If $k = 0$, clearly the predicate reduces to the truth predicate; if $k = 1$, T_1 is identified with non-extensional membership or predicate application. The formulation of the naive schema for T_k is, for arbitrary φ of $\mathcal{L}^+$ with the free variables shown:

(20) $T_k(x_1, \ldots, x_k, \varphi(\hat{u}_1, \ldots, \hat{u}_k, y_1 \ldots, y_n)) \leftrightarrow \varphi(x_1, \ldots, x_k, y_1 \ldots, y_n))$

Here $\varphi(\hat{u}_1, \ldots, \hat{u}_k, y_1 \ldots, y_n)$ stands for a term depending on $\ulcorner\varphi\urcorner$ and the parameters $y_1 \ldots, y_n$, which plays the role of an abstraction term defined by φ and uniformly depending on the parameters shown.

Since, for the sake of consistency, the T_k's have to be partial, the paper outlines a general framework for partial semantics and fixed point techniques, where the natural objects are partial structures, obtained by adding partial predicates for interpreting the T_k-predicates.[207]

The essence of the constructions by Fitch, Gilmore, Kripke, etc., already described in the previous sections (cf. §4.4, §5, etc.), is condensed in the so-called *Fixed-Model Theorem*:

($\diamond$) *Let $\mathcal{L}^+$ be the corresponding language including the predicates $T_0, T_1, \ldots$ and assume that it has a monotone semantics. For any model $\mathcal{M}_0$ satisfying S_0, there is the (least) partial structure $M^* := \langle \mathcal{M}_0, \ldots, T_k^M, \ldots \rangle$, such that, for*

[207] One considers expansions of the form $\langle \mathcal{M}_0, R \rangle$, where R is a three-valued function R of the appropriate arity, or a disjoint pairs of k-ary relations on the universe M_0 of $\mathcal{M}_0$; see §4.4, §5, §5.1, §5.2). Partial structures are naturally ordered according to the relation induced by the poset on the space of truth-values $\mathbf{3} = \{t, f, u\}$.

each k, and each $\varphi(\hat{u}_1, \ldots, \hat{u}_k, y_1 \ldots, y_n)$ in $\mathcal{L}^+$, then

$$\|T_k(a_1, \ldots, a_k, \varphi(\hat{u}_1, \ldots, \hat{u}_k, b_1 \ldots, b_n)\| = \|\varphi(a_1, \ldots, a_k, b_1 \ldots, b_n)\|$$

here above $a_1, \ldots, a_k, b_1 \ldots, b_n$ are (constants naming) arbitrary elements of M; $\|\varphi\|$ is the truth value of the sentence φ according to the chosen monotone semantics.

In order to rephrase the semantical construction in an ordinary axiomatic style, [Feferman, 1984] adopts the intensional solution of [Aczel and Feferman, 1980] and proposes a *non-strict* variant of their $\equiv$-logic, where truth T and falsity F behave like in the Kleene strong three-valued logic and the crucial axiom $T(A \equiv B) \leftrightarrow (A \equiv B)$ is modified to

(21) $T(A \equiv B) \leftrightarrow DA \wedge DB \wedge (A \equiv B)$

The consistency of the schema

(22) $T_k(x_1, \ldots, x_k, \varphi(\hat{u}_1, \ldots, \hat{u}_k, y_1 \ldots, y_n)) \equiv \varphi(x_1, \ldots, x_k, y_1 \ldots, y_n)$,

for arbitrary formulas φ, is implied by a conservation theorem. When (21) is assumed, there is no need to use a Church-Rosser argument, but it is possible to set up a semantics for arbitrary formulas which combines the Fixed-Model Theorem with *two different interpretations of the biconditional*, corresponding to Kleene's logic and to Łukasiewicz's three-valued system.

Let us briefly recall the basic idea, due to Aczel. The informal claim of (22) is that both expressions must have the same truth value (classical or not). So it seems natural to interpret $A \equiv B$ as a Łukasiewicz 3-valued equivalence, which is true whenever A and B have the same truth value, it is undefined if a member is defined and the other is undefined, and it is false in all remaining cases. The problem is that $\|\varphi\|$ is not monotone (with respect to the usual ordering of the truth values) and this forbids an application of the Fixed Model Theorem. The difficulty is avoided by reinterpreting $\equiv$ with Kleene's biconditional . Formally, if φ is an arbitrary formula in the full language (with $\equiv$, $\in$), let φ^K be the formula which one obtains by replacing each occurrence of $\equiv$ with the corresponding Kleene equivalence $\leftrightarrow^k$.[208] Then, of course, the K-transform corresponds to a reinterpretation of the logical operations of φ, which makes $\|\varphi^K\|$ *monotone*. Hence by the "Fixed-Model Theorem", one can find a partial structure satisfying

(23) $\|T_k(a_1, \ldots, a_k, \varphi(\hat{u}_1, \ldots, \hat{u}_k, b_1 \ldots, b_n)\| = \|\varphi^K(a_1, \ldots, a_k, b_1 \ldots, b_n)\|$

Finally, if we consider the two-valued structure $\mathcal{M}$, consisting of the expansion of $\mathcal{M}_0$ with the positive part of the partial predicates solving (23), the classical relation $\mathcal{M} \models \varphi$ can be extended to arbitrary formulas of $\mathcal{L}^+$ by

$$\mathcal{M} \models \varphi \equiv \psi \;\Leftrightarrow\; \|\varphi^K\| = \|\psi^K\|$$

[208] It is important to note that $a \leftrightarrow^k b$ is true if a, b are both true or false; it is false if one is true and the other is false; it is undefined elsewhere, so also if both are undefined.

It is then possible to check that the axioms on $\equiv$ together with (22) hold and there is a common core of the two $\equiv$-variants, which can be formalized as a (non-normal) modal logic.

The final conclusion of [Feferman, 1984] is rather problematic and it is noticed that the mathematical content of the theories considered is rather meager. The alternative seems instead to develop a *theory of total classes* and *partial functions*. This is just the route followed with the program of explicit mathematics we briefly touched upon in section 4.4

5.5 From naive semantics to general semi-inductive processes

At the beginning of the eighties, A. Gupta [1982] and H. Herzberger [1982b] independently discover that semantical paradoxes can be treated by means of *the classical valuation schema*.[209] The underlying motivation is that paradoxes arise *in the context of classical logic and they ought to be solved in accordance with the same logic*.[210] Thus paradoxical sentences are not neutralized by regarding them as both true and false (there exist truth values gluts) or neither true nor false (there exist gaps); on the contrary, a sentence can only receive a classical truth value. However, like in Kripke's theory, semantical evaluation takes place *"along a process f"*, and hence the fundamental characteristic of a sentence is that it is naturally endowed with a *valuational pattern*.

A crucial feature of this approach, originally emphasized by Gupta [1982], has a conceptual flavour: according to him, the concept of truth is characterized by a *revision procedure*. So "when we learn the meaning of 'truth', what we learn is a rule that enables us to improve on a proposed candidate for the extension of truth"([Gupta, 1982]). In dealing with truth, we meet *an inherently circular notion*, whose denotation we can try to fix only by proposing a guess X, then by refining X to a better candidate $J(X)$, and so on, until we eventually get some definite value (if any, or at least a general pattern of instability).

Following Herzberger [1982b], let us briefly inquire what happens if we start out Kripke's evaluation procedure having the classical evaluation method. If $\mathbb{N}$ is the standard model of PA, define

(24) $J(X) = \{\ulcorner\varphi\urcorner, \mid \langle \mathbb{N}, X\rangle \models \varphi\}$

and the iteration

$$J^X(0) = \emptyset \tag{25}$$

$$J^X(n+1) = J(J^X(n)). \tag{26}$$

Since there are no gaps and the evaluation is non-monotone, no fixed point can be reached. For instance, if $\langle \mathbb{N}, S\rangle \models L$, where S is a subset of natural numbers

[209] Below we emphasize the specific technical aspects, without entering the philosophical discussions in which both papers are rooted.

[210] This tenet might be considered as qualifying the approach as "naive", at least in Herzberger's view.

interpreting the truth predicate T and L is a Liar sentence satisfying $L \leftrightarrow \neg T(\lceil L \rceil)$, then

$$(\mathbb{N}, J(S)) \models T(\lceil L \rceil), \text{ i.e. } (\mathbb{N}, J(S)) \models \neg L;$$

hence by iterating the procedure one has, successively:

$$(\mathbb{N}, J^2(S)) \models L, \quad (\mathbb{N}, J^3(S)) \models \neg L, \quad (\mathbb{N}, J^4(S)) \models L, \ldots$$

However, the process does extend into the transfinite by using the operation of $\liminf$ (lower limit). If λ is a limit and f is an operation from ordinals $< \lambda$ to subsets of $\mathbb{N}$:

$$\liminf_{\alpha<\lambda} f(\alpha) = \{x \mid (\exists \delta < \lambda)(\forall \gamma)(\delta \leq \gamma < \lambda \rightarrow x \in f(\gamma))\}; \tag{27}$$

$$\liminf_{\alpha<\lambda} f^-(\alpha) = \{x \mid (\exists \delta < \lambda)(\forall \gamma)(\delta \leq \gamma < \lambda \rightarrow x \notin f(\gamma))\} \tag{28}$$

Then we can define:

$$J^X(\lambda) = \liminf_{\alpha<\lambda} J^X(\alpha) \tag{29}$$

As noticed by Herzberger, the construction defined by the clauses (25), (26), (29) is better understood in a general setting, where the role of the semantical evaluation schema embodied by the operator of (24) is replaced by any "jump operator" $J : \mathcal{P}(M) \rightarrow \mathcal{P}(M)$, mapping the powerset of M (the so-called "base set") into itself. So a *semi-inductive process* in the sense of Herzberger (henceforth a H-process, in short) is any transfinite ordinal sequence $f(\alpha)$ satisfying

$$\begin{aligned} f(0) &\in \mathcal{P}(M) \\ f(\alpha+1) &= J(f(\alpha)) \\ f(\lambda) &= \liminf_{\alpha<\lambda} f(\alpha) \end{aligned}$$

Therefore, to every H-process we can naturally associate:

(i) the set of *positively stable* elements

$$S_f^+ = \{a : (\exists \beta)(\forall \delta \geq \beta)(a \in f(\delta))\}; \tag{30}$$

(ii) the set of *negatively stable* elements

$$S_f^- = \{a : (\exists \beta)(\forall \delta \geq \beta)(a \notin f(\delta))\}; \tag{31}$$

(iii) the set $S_f := S_f^+ \cup S_f^-$ of *stable* elements and its complement, the set of *unstable* elements.

It turns out that H-processes share important properties,[211] namely:

[211] As to the proofs, which use essentially standard set theoretic methods, they are not given in the published version of [Herzberger, 1982b], but they can be found in subsequent work by Burgess [1986], McGee [1991], Visser [1989].

1. for every H-process f, there exists a "stabilization ordinal" κ [212] such that

   (32) $S_f^+ = f(\kappa) = \liminf_{\alpha<\kappa} f(\alpha)$ and $S_f^- = \liminf_{\alpha<\kappa} f^-(\alpha)$

2. every H-process f is *almost everywhere periodic*, that is, there exist ordinals κ and ρ (*the period*) such that, for every ν and every $\xi > \kappa$,

   $$f(\kappa) = f(\kappa + \rho \cdot \nu); \tag{33}$$
   $$f(\xi) = f(\kappa + \mu); \tag{34}$$

   for some $\mu < \rho$.

3. for any H-process f, there exists an ordinal (*saturation point*) μ such that, for all α, there exists $\gamma < \mu$ such that $f(\alpha) = f(\gamma)$.

By (33), (34), "once a saturation point is reached, every subsequent stage repeats some earlier stage and the process becomes cyclic". So the process can be decomposed into an initial piece till the stabilization point κ is reached; then the process enters a cycle of a fixed length, the period ρ, which repeats endlessly all terms $f(\beta)$ of the sequence, for $\kappa \leq \beta < \kappa + \rho$ (the so called *grand loop*).

In a certain sense, the results above show that the Herzberger semi-inductive processes are adequate for the intended semantical application. In fact, the H-process induced by the classical valuation schema exactly generates at a given stage κ exactly the set of sentences S_f^+ (S_f^+), which yields an interpretation of truth (falsehood) as the set of stable truths (the set of stable falsehoods). *The initial piece of the process has the effect of filtering out the unstable sentences*, so that at stage κ exactly the sentences with a stable truth value are generated. Then unstable sentences enter the process again, but after ρ steps they are again filtered out.

As a criticism of H-processes, one can underline their dependence upon the initial points (the assignment to the truth predicate T at stage 0) and also the fact that the definition of the limit rule is clearly somewhat arbitrary. Gupta [1982] and Belnap [1982] have proposed variants.

Typically, one may stick to the intuition that the initial guess (the value of $f(0)$) deserves a priority: so, if a sentence is unstable below λ, then it is made true or false according to the value it receives from the initial guess. This is the idea behind Gupta's limit rule:

(35) $f(\lambda) = \liminf_{\alpha<\lambda} f(\alpha) \cup (f(0) - \liminf_{\alpha<\kappa} f^-(\alpha))$

Belnap [1982] adopts a more general solution, which is inspired by the idea that a revision process should rely *only* upon the jump operation (e.g. on Tarski's valuation schema) and the arbitrariness of the initial guess and of the limit choices

[212] This also called *enclosure ordinal* in [Herzberger, 1982b] and *alignment point* in [Herzberger, 1982a]. Clearly κ will depend on the jump f and its underlying structure. That it cannot made independent on the initial choice $f(0)$, is shown in [McGee, 1991], p. 134.

ought to be eliminated by quantifying on all possible "bootstrapping policies". A B-process (i.e. a semi-inductive process in the sense of Belnap) is then characterized by any limit clause satisfying

(36) $\liminf_{\alpha<\lambda} f(\alpha) \subseteq f(\lambda) \subseteq -\liminf_{\alpha<\lambda} f^{-}(\alpha)$.

It should be noticed at this stage that the properties stated for H-processes *do not hold for arbitrary semi-inductive processes*; yet, it is possible to prove suitable generalizations thereof, following [Burgess, 1986] and [McGee, 1991]. So, given a process f (where we can assume that each $f(\alpha)$ is a subset of natural numbers), we define an ordinal κ *reflecting for* f provided it meets the conditions:

$$S_f^{+} = \liminf_{\alpha<\kappa} f(\alpha) \text{ and } S_f^{-} = \liminf_{\alpha<\kappa} f^{-}(\alpha) \tag{37}$$

$$\liminf_{\alpha<\kappa} f(\alpha) \subseteq f(\eta) \subseteq -\liminf_{\alpha<\kappa} f^{-}(\alpha) \tag{38}$$

for all $\kappa < \eta$. Then it can be shown:

(*) The set of all limit ordinals ξ that are reflecting for f contains a closed and unbounded subset of ordinals.[213]

For further developments towards a general theory of revision sequences, the reader should consider [Gupta and Belnap, 1993], ch.5, 5C.

We cannot conclude our brief historical survey on the birth of the revision theory without mentioning that the semi-inductive approach to semantics yields a rather subtle method for classifying self-referential statements, which is as powerful as Kripke's. For the sake of comparison with Kripke's, we only recall a few examples, following the notation of [Burgess, 1986].

Let $P := H, G, B;$, we assign a sentence φ *a P-type* S*,* where S is any label from the list $\langle T, F, U, TF, TU, FU, TFU\rangle$, and $T(F, U)$ occurs in S if and only if φ is stably true (stably false, unstable) for some P-process.[214] So, while the Liar is of type U over all processes, every logically valid sentence, by contrast with the Kripkean case, is always of type T. It is easy to check that the Truth Teller is of type TF for each process, while the conjunction χ of a Truth Teller with the disjunction of a Truth Teller and a Liar is of type TFU. Moreover, there are concrete examples of situations, which turn to be ungrounded on Kripke's approach, which can be made easily semantically definite by means of the classical valuation.[215]

In the present context, one can even distinguish several paradoxical sentences by explicitly taking into account another crucial parameter, *the degree of instability*, which can be assigned to a given sentence, as a measure of its cyclic valuation

[213] A set X of ordinals is *unbounded* if for every ordinal α, there is an ordinal $\beta \in X$ such that $\alpha < \beta$; X is *closed* if and only if every bounded subset of X has a sup in X.

[214] If a sentence is of H-type TFU, then there are (at least three distinct) H-processes where the sentence is stably true, stably false and unstable (respectively).

[215] For Gupta's puzzle, see [Gupta, 1982], pp. 34-35.

pattern. For instance, the Liar is paradoxical and oscillating with the pattern $1, 0, 1, 0, \ldots$, and hence it has period 2. $BL := TL \vee T\neg L$ is paradoxical as well, but it has period ω;[216] indeed, one can produce paradoxical statements with large transfinite cycles.

A closer look clearly shows that "revision theories of truth are complicated" in the technical sense of generalized recursion theory and this naturally raises "definability issues" ([Welch, 2001a]). Let us only present a concise account of the central results, which are known up to now.

We stick to the notation from Burgess [1986]. If $P := H(G, B)$, $\Box_P T$ ($\Box_P F$, $\Box_P U$, $\Diamond_P T$, $\Diamond_P F$, $\Diamond_P U$) denotes the class of all (Gödel-numbers of) sentences φ, such that φ is stably true (stably false, unstable) over all (some) P-processes.[217] Burgess [1986] and McGee [1991] have shown

(†) *the sets* $\Box_H T$, $\Box_G T$, $\Box_B T$ *are all* Π^1_2*-complete.*

The result has then been extended to the general framework of the *theory of arbitrary circular definitions* in [Kremer, 1993], [Antonelli, 1994c], [Antonelli, 1998] -[Antonelli, 1994b] and quite recently by [Welch, 2001a], [Welch, 2001b], [Löwe and Welch, 2000].

What is a circular concept G? Well, G is simply a notion that must satisfy a circular condition, just like the Tarski biconditional $T(\lceil\varphi\rceil) \leftrightarrow \varphi$, where φ may contain T itself.

In order to make this precise, Gupta and Belnap consider any first order language $\mathcal{L}$ (think of the standard arithmetical language) and its expansion $\mathcal{L}^+$ with possibly denumerably many predicate symbols $G_n(x_1, \ldots, x_n)$.

A set $\mathcal{D}$ of circular definitions is simply a set of conditions of the form

(39) $G(\vec{a}) =_{df} A_G(\vec{a})$

where $A_G(\vec{a})$ is formula of $\mathcal{L}^+$, possibly containing some G_n's.

In analogy with the case of truth, the idea is that, in order to assign a meaning to G with respect to a given structure $\mathcal{M}$ for $\mathcal{L}$, a revision process must be invoked, where we start by suggesting an *initial hypothesis* h[218] on the extension of G, and we keep revising it by means of a jump operation, usually embodied by a formula $A^G(\vec{a})$. Now, given $\mathcal{M}$, a Γ-process f is simply a revision process, where we have adopted a definite policy (or rule) Γ at limit stages and an initial hypothesis h. This leads to a suitable general notion of definability:

(i) a set $X \subseteq |\mathcal{M}|$ is *weakly definable* in S_Γ if and only if there exists a formula $\varphi(x)$ of $\mathcal{L}^+$ which defines X in S_Γ, i.e. for all $a \in |\mathcal{M}|$, $a \in X$ if and only if $\varphi[a]$ is stably true in every Γ-process.

[216] In fact, BL is false at stage 0, but true at any finite subsequent stage; thereafter it is false at stage ω, but again true at every stage of the form $\omega + k$ (k finite), etc.

[217] So $\Diamond_G F$ is the class of all (Gödel-numbers of) sentences φ, such that φ is stably false over some G-process.

[218] h is a subset of $|\mathcal{M}|^n$, n being the arity of the circular predicate G.

(ii) a set $X \subseteq |\mathcal{M}|$ is *strongly definable* in S_Γ if and only if it is weakly definable in S_Γ by means of a formula $\varphi(x)$, and its negation weakly defines the complement of X in S_Γ.

If we identify $\mathcal{M}$ with $\mathbb{N}$, the problem is: which sets of natural numbers are definable when the complexity of the A_G's and Γ varies? The main theorem is due to Welch [2001a]:

(♮) *if* Γ *is* Δ^1_2*-definable ,*[219] *then the class of* S_Γ*-definable reals (where 'real' means 'set of natural numbers') is precisely the class of* Π^1_2*-reals*

(♯) *on the same assumption, the class of* Δ^1_2*-reals and the class of strongly* S_Γ *definable reals coincide.*

As noticed by Welch, *a nice pattern emerges, linking formal semantics, generalized recursion theory and admissible set theory.* On one side, we know that there are deep connections between the Kripke theory of truth and the theory of inductive definitions, and hence the notion of *next admissible set.* In fact, given an arbitrary acceptable first order structure $\mathcal{M}$, the set of inductive subsets over $\mathcal{M}$ coincide with the collection of sets which are Σ_1-definable over the next admissible structure $HYP_\mathcal{M}$.[220] On the other side, revision theoretic definability is deeply related to the notion of "next stable structure" $\mathbb{S}_\mathcal{M}$:[221]

(♭) *let* $\mathcal{M}$ *be an acceptable structure in the sense of Moschovakis [1974], and let* Γ *be* Δ_1*-definable over the universe of hereditarily countable sets generated above* $\mathcal{M}$*. Then the class of* S_Γ*-definable subsets of* $\mathcal{M}$ *is precisely the class of* $\Pi_1(\mathbb{S}_\mathcal{M})$ *sets.*

(♯) *On the same assumptions, the class of* $\Delta_1(\mathbb{S}_\mathcal{M})$*-subsets of* $|\mathcal{M}|$ *and the class of strongly* S_Γ *definable subsets of* $|\mathcal{M}|$ *coincide.*

REMARK 4.

(i) Apart from analytic inspiration, some germs of revision theoretic ideas can be found in the recursion theory of the fifties. Let us only mention the characterization of Δ^0_2-subsets of natural numbers as *trial and error* predicates, due to Shoenfield (see [Odifreddi, 1989], p. 373).

(ii) Magari [1974], [1975] (English abstract in [Magari, 1973]) has shown that there exists a notion of (experimental) arithmetical truth, which corresponds to recursion theoretic complexity Σ^0_2, see also [Ursini, 1977], [Ursini, 1978];

[219] Upon coding, and this includes the usual H-, G-, B-processes.

[220] Roughly, the least constructible universe, containing $\mathcal{M}$ as an element and satisfying the Kripke-Platek axioms; see [Barwise, 1975].

[221] Roughly, $\mathbb{S}_\mathcal{M}$ is the least constructible universe built above $\mathcal{M}$, such that it is a Σ_1-elementary substructure F the whole universe V, see [Barwise, 1975] for details.

(iii) [Antonelli, 1994c] has indeed applied these tools back to standard definability; e.g. he shows that every set in the arithmetical hierarchy can be given an explicit revision theoretic definition.

(iv) Revision theoretic ideas have also been applied in the construction of non-well-founded sets again by [Antonelli, 1994a].

(v) An ultraproduct construction for defining a notion of self-referential notion of truth is given in [Negri, 1989]; 'true in a ultraproduct' roughly amounts to 'true almost everywhere' and this is somewhat reminiscent of the revision theoretic approach.

5.6 Non-well-founded sets and semantical paradoxes

In 1987 Barwise and Etchemendy propose a novel approach to the Liar paradox and to the semantics of natural languages, which takes place within *a set-theoretic universe, where self-reference and circularity are intrinsic features of propositions and of the states of affairs they describe.* According to [Barwise and Etchemendy, 1987], propositions are *not* linguistic entities and truth is not a predicate of sentences, like in the Tarskian or Kripkean theory. A *state of affairs* corresponds to a structured set theoretic object, whose constituents are respectively a relation R, a list $\vec{a}$ of objects whose length is equal to the arity of R and a *polarity* $i \in \{0, 1\}$, selecting *positive* or *negative* states.[222] Collections of states of affairs, which are called *situations*, play in the present framework the role of partial models or "coordinate systems" with respect to which propositions are evaluated.

It is important to stress that, according to [Barwise and Etchemendy, 1987], a state-of-affairs can involve semantics as well, e.g. it can have the form $\langle Tr, p, i\rangle$, where $i = 0$ or $i = 1$, Tr stands for the property of *being true* and p is a proposition. Moreover, it is allowed to have *propositions that contain themselves as constituents* (so we do have *direct self-reference*). It then follows that an adequate modelling of semantics is only possible insofar as *the universe of sets is not the standard cumulative hierarchy*, but it satisfies the anti-foundation axiom AFA ([Forti and Honsell, 1983], [Aczel, 1988]):

($\diamondsuit$) *given any structure* $\langle a, r\rangle$ *with* $r \subseteq a^2$*, there exists a unique surjective homomorphism* $f : \langle a, r\rangle \mapsto \langle b, \in\rangle$*, for some transitive* b, satisfying, for $x \in a$:

$$f(x) = \{f(y) | \langle y, x\rangle \in r\}.^{223}$$

AFA grants that the set theoretic universe is inhabited by plenty of circular or reflexive[224] objects, satisfying systems of fixed point conditions $\tau(x) = x$, where

[222] Of course, the positive (negative) polarity corresponds to the fact that the given relation obtains (does not obtain) among the objects of the given list. States of affairs which only differ by their polarity are naturally named *dual*.

[223] AFA admits of a geometrical reading, ensuring the existence and uniqueness of decorations for arbitrary graphs ($\langle a, r\rangle$ is any graph, and the homomorphism plays the role of a decoration).

[224] Call x *reflexive* if $x \in x$.

$x \in X$ and X is a set of 'indeterminates'. Roughly, $\tau(x)$ denotes any set built up from indeterminates in X and possibly *atoms* from a given set A, using set theoretic operations, and the import of the anti-foundation axiom essentially amounts to show that the equation has *a unique solution*. For instance, under AFA, there exists a *unique* set Ω satisfying $\{x\} = x$, as well as there are *unique* propositions L, S corresponding to the Liar and to the Samesayer, i.e. satisfying the identities $p = [Fa(p)]$ and $q = [Tr(q)]$.

AFA-universes[225] - whose elements are usually termed *hypersets* - are applied in order to give two different versions of self-referential truth, and hence different responses to semantical paradoxes: the former inspired by Russell and the latter by Austin. It is impossible to give justice to [Barwise and Etchemendy, 1987] within a short space; so we only limit ourselves to a few sketchy hints.

First of all, in either case the language implements direct self-reference by including explicitly demonstratives, like `this`, `that_i` ($i \in \omega$), and a scope operator $\downarrow$. The idea is that if φ is a sentence in which a demonstrative `this` occurs, then `this` in $\downarrow \varphi$ refers to the whole proposition expressed by φ. So to each sentence φ of the language involving a given list `this`, `that_1`, etc., is assigned a unique proposition $Val(\varphi)(p, \overrightarrow{q})$, depending on the parameters p, $\overrightarrow{q}$; moreover, $Val(\downarrow \varphi)(\overrightarrow{q})$ is the unique p solving

$$Val(\downarrow \varphi)(p, \overrightarrow{q}) = p$$

According to the Russellian option, propositions are claims about a model (of the world) M, which is conceived as a collection of states of affairs. A proposition p is *made true (false)* by M - in symbols $M \models p$ ($M \not\models p$) – if and only if for some (no) $s \subseteq M$, $s \models p$ ($s \not\models p$). It is important to notice that, by contrast with the Kripkean approach, the semantical notion of *proposition* and the classical relation $s \models p$ – though relatively standard as to the closure conditions they satisfy – must be introduced as largest fixed points of suitable monotone operators, and this is a typical feature of constructions taking place in AFA-universes.

As to the Liar proposition, it is not made true by any given model, but this very claim cannot become a fact of the model. Thus it is essential to distinguish between *internal negation* and *external denial* of a proposition: the negation establishes a claim about how things are *in the world*, while the denial is just the rejection of an assertion. Therefore, that *p is true (false) in a model M* means that the state of affairs $\langle Tr, p, 1\rangle$ (respectively $\langle Tr, p, 0\rangle$) is an element of M. This should be contrasted with the external fact that $M \models p$ ($M \not\models p$).

This is reminiscent of similar distinctions arising from the Kripke approach and other truth-gap solutions. In spite of the use of non-well-founded and co-inductive objects, inductive methods à la Kripke are still essential for building up *almost semantically closed models*,[226] which correspond to the best possible approximations to semantically closed worlds. Here a model M is almost semantically closed if M is coherent (i.e. no state of affairs with its dual belongs to M), and it holds:

[225] These can be given a category-theoretic characterization as final coalgebras of the power-set functor *Pow*, i.e. roughly as largest fixed points of *Pow*.

[226] Cf. theorem 10, p. 82,

(i) for every proposition p, p is true in M if and only if p is made true by M; (ii) p is false in M if and only if its negation $\bar{p}$ is made true by M.[227] Once more, the lesson from the paradoxes is that we cannot in general find models where we can jump from the assertion that M does not make p true to the fact that p is false in M.

According to the Austinian approach, a sentence does not uniquely determine a proposition, but it is necessary to explicitly fix its context sensitive elements. Thus a proposition properly amounts to a pair of constituents, given by (i) a *situation parameter* s[228] and (ii) a *type* T, which is uniquely fixed by a state of affairs σ. Now a given proposition is true if and only if the situation the proposition is about is of the given type, i.e. $\sigma \in s$.

As to the paradoxes, AFA implies that, for every situation s, there are Liar-propositions f_s claiming that the falsity of f_s is a fact about s. In general, semantical contradictions are avoided *by changes in the situation parameters*. Indeed, when one uses a Liar sentence λ of the form $\neg$`True(this)` to make a statement about the actual situation s_0, it expresses a *false* proposition f_{s_0}; but this very fact can be used to define a larger situation s_1 with respect to which the Liar sentence now expresses a *true* proposition f_{s_0}. Of course, there exists a unique Liar proposition f_{s_1}; again, this is false with respect to s_1 and by iteration, we are led to s_2 verifying f_{s_1}, etc. Thus the Liar paradox "does not force one into abandoning the intuitive idea that the world is total and that any claim is true or it is not true. What it does, is that it forces one to be extraordinary sensitive to subtle shifts in context" [Barwise and Moss, 1996, p. 189].

To sum up, the solution of Barwise and Etchemendy is dependent upon an idea of set, which is heavily *anti-hierarchical*: sets are abstractions from arbitrary relational structures or graphs. This is clearly against the traditional type-theoretical cumulative view of sets. An important technical feature of this approach to semantics is that it emphasizes the use of *coinduction* and *corecursion* principles (e.g. the choice of greatest fixed points of monotone operators, bisimulation as a way of comparing observationally equivalent systems) and hence it is in some sense dual to the inductive methods emphasized in previous proposals.

The Barwise-Etchemendy semantical theory is part of a general methodology for dealing with the mathematics of non-well-founded phenomena, which since recent years is prompted by several developments coming from linguistics, computer science, philosophy, cognitive science.[229]

REMARK 5. The Austinian theory is an instance of *the contextual approach,* that is, of a solution to the paradoxes which takes into account the context of utterance of a sentence and attempts to render these ideas with mathematical tools. Other candidates in this directions are [Parsons, 1974], [Burge, 1979], [Burge, 1982], [Gaifman, 1988], [Gaifman, 1992], [Simmons, 1988], [Simmons, 1993].

[227] The negation $\bar{p}$ being as unique solution of a suitable set of equations by AFA.

[228] Informally, a limited portion of the real world or a possible context.

[229] For a systematic account, the reader, is urged to consult the monograph of Barwise and Moss [1996].

6 AXIOMATIC THEORIES OF TRUTH

As stressed by Sheard [1994, p. 1044] in his guide to truth predicates in the modern era, the problem of semantical closure and self-referential truth can be tackled with the tools of the formal axiomatic method, as "*one of creating not models but consistent theories*". However, it has to be remarked that the motivations behind the axiomatic work are at first mainly *foundational*, even proof-theoretical (and quite apart from philosophical surroundings), and that they reach a finished (published) form only somewhat later than the related semantical constructions. In the subsequent sections, we consider the principal contributions between the end of the eighties and mid 90's.

6.1 Classifying the "logics of truth": Friedman and Sheard

In this direction, a landmark is represented by the contribution of Friedman-Sheard (1987); there the authors " create a list of plausible axioms and rules of inference concerning the truth predicate T, each of which embodies *some aspect of its intended interpretation as truth*". Then they *classify all subsets of the list as either consistent or inconsistent, and they measure the proof-theoretic strength of several subsets by comparing them with familiar systems of arithmetic and analysis*" ([Friedman and Sheard, 1987, p. 1]).

The paper in question marks a point of high sophistication: techniques from recursion theory and proof theory interact with modal intuitions; moreover, certain scattered metamathematical contributions (e.g. see [Kaplan and Montague, 1960], [Montague, 1963]) of the early sixties, which could be regarded as refinements of Tarski's and Gödel's theorems, but also had motivations in philosophical logic,[230] are here systematically developed in a unified framework. The generality of the results contained in [Friedman and Sheard, 1987], [1991] not only provide a useful way to detect possible inconsistencies and consistency results, but also make it reasonable to speak of *logics of truth*.

Let us briefly summarize the core of the principal paper [Friedman and Sheard, 1987]. First of all, the axiomatic investigation is built upon a base theory $Base_T$, which consists of a weak fragment of Peano arithmetic (say, including primitive recursive arithmetic) in the language expanded with the truth predicate, where syntax and a certain amount of semantical notions can be formalized. Thus it is assumed that there is a predicate $Val(x)$ expressing that x is the Gödel number

[230] Let us quote as an example (theorem 3) of Montague [1963]. If T-Out is the schema $T(\ulcorner A \urcorner) \rightarrow A$, then T-Out is inconsistent with the rule T-Intro of T-introduction $A/T(\ulcorner A \urcorner)$ (modulo a ground theory Q which includes enough of arithmetic, to prove the diagonalization theorem). The result can then be regarded as a negative verdict against a syntactical treatment of modality, and in particular against interpreting the verb "to know" as a predicate of sentences (it actually arose in a formalization of the so-called Hangman Paradox and this explains why it was termed Knower's paradox). Just to cite Montague, "if necessity is to be treated syntactically, that is, as a predicate of sentences, as Carnap and Quine have urged, then virtually all of modal logic must be sacrificed"(cf. [Montague, 1963], p. 294). A sharpening of the paradox is later achieved by Thomason [1980].

of a valid first-order formula. Also, if $\circ$ is a logical symbol, there is a definable function symbol $\dot{\circ}$, naming the corresponding arithmetical operation on codes, etc. Now the basic axioms stipulate that T satisfies

$$Val(x) \rightarrow T(x)$$
$$T(x \dot{\rightarrow} y) \rightarrow T(x) \rightarrow T(y)$$

Hence truth is always closed under *classical first order logical consequence* and this is in marked contrast with all approaches stemming from proposals based on gap semantics.

Following Friedman and Sheard, we now recall the list *Optional Axioms*, made up with truth axioms and rules, which form the basic combinatorial ingredients. In order to make clear the intuitive content and possible analogies with *modal principles*, we will abuse notation and we will not distinguish between a formula and its code numbers, leaving to the reader an appropriate strict formalization.[231] As to the *Truth Axioms*, they include:

- T-In: $A \rightarrow TA$;
- T-Out: $TA \rightarrow A$;
- T-Rep: $TA \rightarrow TTA$;
- T-Del: $TTA \rightarrow TA$;
- T-Cons: $\neg(TA \wedge T\neg A)$;
- T-Comp: $TA \vee T\neg A$;
- U-Inf: $\forall n T(A(n)) \rightarrow T((\forall x)A(x))$;
- E-Inf: $T((\exists x)A(x)) \rightarrow \exists n T(A(n))$.

Then four *Truth Rules* are also considered:

- T-Intro: $A \ / \ TA$;
- T-Elim: $TA \ / \ A$;
- $\neg$T-Intro: $\neg A \ / \ \neg TA$;
- $\neg$T-Elim: $\neg TA \ / \ \neg A$.

The main theorem states that *the following are the only maximal consistent subsets of the optional list*:

(A) *T-In, T-Rep,* $\neg$*T-Elim, T-Intro, T-Del, T-Comp, U-Inf, E-inf*;

[231] In particular, when we write $(\forall n)T(A(n))$, we understand that the quantification within the scope of T makes sense because we can use a standard formalized substitution operation.

(B) *T-Rep, T-Cons, T-Comp, U-inf, E-Inf;*

(C) *T-Del, T-Cons, T-Comp, U-Inf, E-Inf;*

(D) $\neg$*T-Elim, T-Intro, T-Elim,* $\neg$*T-Intro, T-Cons, T-Comp, U-Inf, E-Inf;*

(E) *T-Intro, T-Elim,* $\neg$*T-Intro, T-Del, T-Cons, U-Inf;*

(F) $\neg$*T-Elim, T-Intro, T-Elim, T-Del, U-Inf;*

(G) *T-Rep,* $\neg$*T-Elim, T-Intro, T-Elim, U-Inf;*

(H) *T-Out, T-Rep, T-Elim,* $\neg$*T-Intro, T-Del, T-Cons, U-Inf;*

(I) *T-Rep,* $\neg$*T-Elim, T-Elim, T-Del, U-Inf.*

The proof consists of displaying ω-models or else suitable consistent sets satisfying the required optional axioms, and, in order to get maximality, by a systematic recollection of various inconsistencies. Among them, it is worth mentioning the proof of Löb's theorem in the form

(40) $T(TA \to A) \to TA$

from the axioms $Base_T$ with T-Rep and T-Intro.

Connections with proof theory are made apparent by two theorems:

(i) Let FS$\lceil$ be the system including $Base_T$ with the optional axioms $\neg$T-Elim, T-Intro, T-Elim, $\neg$T-Intro, T-Cons, T-Comp. Then FS$\lceil$ is a conservative extension of Peano Arithmetic.

(ii) Let T_0 be $Base_T$ with the optional axioms T-Out, T-Elim, $\neg$T-Intro, T-Del, T-Cons, U-Inf. Then T_0 is conservative over the first order theory ID_1 of one inductive definition (or the subsystem BI of second order arithmetic).

[Friedman and Sheard, 1987] raised the question whether (ii) holds even if T-Rep is added; the problem has been solved in [Cantini, 1990] by proof theoretic methods (and actually by investigating a cut free infinitary calculus based on inferences which are essentially Gentzen-like versions of T-intro and $\neg$T-Intro). In the same paper, it is also shown that the Friedman-Sheard theory is naturally related to the van Fraassen supervaluation semantics.

Recently, Halbach [1994] has shown that the result (i) extends to the case where even U-Inf and E-inf are assumed. If the induction schema is restricted to *arithmetical formulas*, the restricted system FS$\lceil$, consisting of T-intro, T-elim, T-Cons, T-Comp, and the quantifier principles U-Inf and E-Inf, is conservative over Peano Arithmetic. Moreover, if FS is the system which includes *the induction schema for arbitrary formulas*, then it has the same strength as *ramified analysis of rank at most* ω (and it roughly corresponds to a theory of finitely iterated arithmetical truth).

Clearly the systems FS and FS⌈ are very close to axiomatize *a notion of consistent and complete truth*; yet, full adequacy is not reached (as T-Rep is not present, as it would be inconsistent).

On the conceptual side, though the FS is a good approximation to a naive notion of (arithmetical) truth, FS has no ω-model. In fact, in 1987 McGee [1985] proved with an ingenious fixed point argument that *any theory including U-Inf, T-Cons and closed under T-Intro must be ω-inconsistent*.

On ill-founded Tarskian hierarchies

Another example of ω-inconsistent theory of truth is given by Visser [1989]. There he shows the existence of an infinite descending sequence $\{L_n \mid n \in \omega\}$ of Tarskian languages,[232] such that L_0 contains a truth predicate T_0 for L_1, L_1 has a truth predicate T_1 for L_2, etc. The resulting set

$$Th(\mathbb{N}) + \{T_i(\ulcorner\alpha\urcorner) \leftrightarrow \alpha \mid \alpha \text{ sentence of } L_{i+1},\ i \in \omega\}$$

(where $Th(\mathbb{N}$ is the set of sentences true in the standard model $\mathbb{N}$ of PA) is easily seen consistent by compactness. However, already the set $PA + \{T_i(\ulcorner\alpha\urcorner) \leftrightarrow \alpha \mid \alpha \text{ sentence of } L_{i+1},\ i \in \omega\}$ is ω-inconsistent

6.2 Reflective closures

From a wholly different perspective, axiomatic theories of self-referential truth, which formalize the closure conditions related to Kripke's jump operator, were investigated by Feferman in connection with an unramified approach to predicativity and with the general problem of the incompleteness of formal systems (reflective closure of a theory, see [Feferman, 1991]). The idea is to give a formally precise answer to the question: "which statements in the base language L of a theory S ought to be accepted if one has accepted the basic axioms and rules of S ? " The answer receives an axiomatic formalization by means of Feferman's notion of *weak reflective closure Ref(S)* of S, which consists in expanding the ground system with a notion of partial self-applicable truth.

For example, let S:=PA and assume we actually work in a definitional extension of Peano arithmetic, which represents and proves the basic fact about elementary syntax[233] and whose language includes two additional unary predicate T and F .

[232] An apparently independent paper is [McCarthy, 1988]. By the way, there is a hint already in Kripke [1975].

[233] Of course, we presuppose a standard encoding of the ground arithmetical language $\mathcal{L}$: $\#E$ is the code of E (E being an expression, i.e. a term or a formula of $\mathcal{L}$); $\ulcorner E\urcorner$ stands for the term (actually the numeral) representing $\#E$. So $CT(x)$, $For(x)$, $Sent(x)$ are (respectively) the predicates "x is the code of a closed $\mathcal{L}$-term", "x is the code of a formula ", "x is the code of a sentence"; if O is a logical symbol, the dot notation $\dot{O}$ stands for the corresponding (syntactical representation of the same) operations defined on codes of terms, formulas, etc. The notation $a(x)$ is a shortening for the term representing the operation of substituting the first free variable of the expression encoded by a with the x*th*-numeral. $val(x)$ designates the operation, associating to the term encoded by x the numeral representing its value.

Then *Ref(PA)* is the theory which includes Peano axioms with *the induction schema extended to the whole language* and the so-called *Self-Truth axioms*:

(i) $CT(x) \wedge CT(y) \rightarrow ((T(x \doteq y) \leftrightarrow val(x) = val(y)) \wedge (F(x \doteq y) \leftrightarrow \neg val(x) = val(y)))$;

(ii) $CT(x) \rightarrow (T(\dot{T}(x)) \leftrightarrow T(val(x))) \wedge (F(\dot{T}(x)) \leftrightarrow F(val(x)))$;

(iii) $CT(x) \rightarrow (T(\dot{F}(x)) \leftrightarrow F(val(x))) \wedge (F(\dot{F}(x)) \leftrightarrow T(val(x)))$;

(iv) $(T(\dot{\neg}(x)) \leftrightarrow F(x)) \wedge (F(\dot{\neg}(x) \leftrightarrow T(x)))$;

(v) $(T(x) \wedge T(y) \leftrightarrow T(x \dot{\wedge} y)) \wedge (F(x) \vee F(y) \leftrightarrow F(x \dot{\wedge} y))$;

(vi) $((\forall x)T(a(x)) \leftrightarrow T(\dot{\forall} a)) \wedge ((\exists x)F(a(x)) \leftrightarrow F(\dot{\forall} a))$.

The main theorem states

$$Ref(PA) \equiv RA_{<\epsilon_0}$$

This means that the ordinary reflective closure of arithmetic is in fact equivalent to predicative analysis of any level up to the first ϵ-number ϵ_0 i.e. roughly to the fragment of second order arithmetic, which is based upon transfinite iteration (along ϵ_0) of the standard arithmetical comprehension

$$(\exists X)(\forall u)(u \in X \leftrightarrow \varphi(u))$$

φ being a formula containing only number theoretic quantifiers and possibly free second order variables.

The notion of ordinary reflective closure can be also extended by a suitable substitution rule in order to answer the problem of which *schemata* can be regarded as implicit in accepting a given list of *schematic axioms and rules*. It turns out that the resulting notion of *schematic reflective closure* $Ref^*(PA(P))$ yields an alternative unramified characterization of predicative analysis (cf. [Feferman, 2005]:

$$Ref^*(PA(P)) \equiv RA_{<\Gamma_0}{}^{234}$$

The example witnesses how investigations deriving from formal semantics and paradoxes have reached a high level of integration with different advanced areas of mathematical logic; for instance, while the lower bound result is actually a refinement of typical predicative well-ordering techniques, the upper bound theorem exploits sophisticated methods and results from reductive proof theory, depending on the fact that $Ref(PA)$ can be modelled in fixed point theories, and that these are interpretable, via a trick of Aczel [1977], in the subsystem $\Sigma^1_1 - AC$ of second order arithmetic, which is known, since a celebrated result of Friedman (1968), to be predicatively reducible to $RA_{<\epsilon_0}$. It is worth recalling that this proof of

[234] Γ_0 is the first strongly critical ordinal, which is known to be the first impredicative ordinal (in the sense of the Feferman-Schütte characterization of predicativity).

the upper bound result uses the fact that $Ref(PA)$ does not exclude non-minimal fixed point models,[235] since the consistency axiom Cons

$$(\forall x)(Sent(x) \rightarrow (\neg T(x) \vee \neg F(x))$$

is not assumed.

REMARK 6. That the main theorem still holds once consistency is assumed, has been proved in [Cantini, 1989]. Furthermore, the system where *consistency* is replaced by *completeness*, i.e. by the axiom

$$(\forall x)(Sent(x) \rightarrow (T(x) \vee F(x))$$

has the same strength (it is actually intertranslatable with the theory based on consistent truth) and it corresponds to axiomatizing the largest fixed point of the Kripkean operator of §5.3.

It turns out that *suitable variants of 'reflective closure' can be profitably applied as tools for proof-theoretic investigations*: for instance, as an intermediate step for computing the proof theoretic strength of transfinitely iterated fixed point theories (see [Jäger *et al.*, 1999]); or for designing strong systems conservative over PA or even weaker systems (see [Cantini, 1997]).

6.3 Recursively saturated models of PA and classical arithmetical truth

Classical arithmetical truth has been investigated in the context of non-standard models of PA. Indeed, once formal languages are arithmetized, one necessarily meets in arbitrary models of PA *non-standard correlates* of syntactical notions[236] and the question arises: for which models of PA can we define a notion of truth (or satisfaction) for such non-standard languages?

More precisely, if $\mathcal{M}$ is a model of PA, $L(\mathcal{M})$ is the language obtained by adjoining constants naming the elements of $\mathcal{M}$, and $L^*(\mathcal{M})$ is the set of formulas in the sense of $\mathcal{M}$, it is natural to inquire whether there exists a *full satisfaction class* for $\mathcal{M}$, that is, a subset Σ of sentences of $L^*(\mathcal{M})$, containing all true atomic and negated atomic sentences, satisfying the conditions

- for every sentence φ of $L^*(\mathcal{M})$, exactly one of φ and $\neg\varphi$ belongs to Σ;
- for every sentence φ, ψ of $L^*(\mathcal{M})$, $\psi \vee \varphi \in \Sigma$ if and only if at least one of φ, ψ belongs to Σ;
- for each sentence $\exists x\psi$ of $L^*(\mathcal{M})$, $\exists x.\psi \in \Sigma$ if and only if $\psi(a) \in \Sigma$, for some $a \in |\mathcal{M}|$.

[235] So in a sense the theory axiomatizes the closure properties of an arbitrary fixed point model, see §5.2.

[236] This was noticed as early as 1963 by Abraham Robinson [1963].

Since the mid seventies, the problem has received fruitful attention (see [Kotlarski *et al.*, 1981], [Lachlan, 1981], [Smith, 1987], [Kotlarski, 1991], [Kaye, 1991], [Kotlarski and Z.Ratajczyk, 1990]) and a highlight of this research is certainly the discovery of *the connection between recursive saturation*[237] *and existence of satisfaction classes over non-standard countable PA-models*:

(i) Lachlan [1981] showed that, *if a non-standard model $\mathcal{M}$ of PA has a satisfaction class, then $\mathcal{M}$ is recursively saturated*;

(ii) conversely, [Kotlarski *et al.*, 1981] has established that *every countable recursively saturated model of PA has a full satisfaction class.*[238]

The proof of theorem (ii) is ingenious and it is based upon a proof theoretic characterization of recursive saturation by means of $\mathcal{M}$-logic relative to a given countable $\mathcal{M}$; recursive saturation is actually equivalent to the consistency of the empty set (of sentences) in $\mathcal{M}$-logic. Another crucial idea is that of finite approximation of non-standard formulas.

Quite recently, with different motivations, stemming from the philosophical debate on the so-called deflationist theories of truth, Halbach [1994] noticed that the main theorem of [Kotlarski *et al.*, 1981] has an interesting proof theoretic by-product for theories of classical arithmetical truth.

Let AT (acronym for "arithmetical truth") be the natural extension of Peano arithmetic PA, which comprises: 1. the inductive clauses of the classical Tarskian truth predicate T, restricted to (arbitrary codes of) arithmetical (i.e. T-free) formulas; 2. the axiom T-ind, saying that every arithmetical instance of the induction principle is true.

Then *AT and PA have the same arithmetical theorems*;[239] for, if $\mathcal{M}$ is an arbitrary countable model of PA and φ is an arithmetical sentence, then there also exists a recursively saturated elementary (countable) extension $\mathcal{M}^*$, which satisfies φ; by the Kotlarski-Krajewski-Lachlan theorem, there exists a subset X of $\mathcal{M}^*$, having the closure conditions of a satisfaction class and hence also satisfying the axioms of AT. [Halbach, 1999] also shows that, if only arithmetical induction is allowed and T-ind is omitted from AT, then the resulting system is proof-theoretically reducible to PA[240] by means of a suitable finitary cut elimination argument.

[237] Recall that a structure $\mathcal{M}$ is *recursively saturated* if and only if, for every $n \in \omega$, for every *recursive* set $\Phi(x_1, \ldots, x_n)$ of $L(\mathcal{M})$-formulas with free variables in the list $x_1, \ldots, x_n$, if every finite subset of $\Phi(x_1, \ldots, x_n)$ is satisfiable in $\mathcal{M}$, then $\Phi(x_1, \ldots, x_n)$ is satisfiable in $\mathcal{M}$.

[238] The satisfaction class may be chosen to be highly pathological, i.e. to contain the negation of a non standard sentence which is a disjunction of $0 = 0$.

[239] Feferman [1991] has considered (a version of) AT without T-ind, but endowed with arithmetical induction, under the name of "ordinary truth theory of PA", and he has stated its conservativity over PA.

[240] Of course AT with *full induction schema* is much stronger than AT and can be proven of the same strength as arithmetical comprehension with full induction.

7 CONCLUSION

This final section has a three-fold aim:

(i) it quickly surveys additional logical results, related to paradoxes, truth, self-reference, which have been published in the interval 1990-2002;

(ii) it considers a few tentative applications in different fields (Theoretical Linguistics, Computer Science, AI);

(iii) it fills up some gaps in the previous exposition.

We do not claim completeness and the choice largely depends on the interests of the present author; we simply wish to complete the picture with a sample of updated references and an idea of future directions.[241]

7.1 Miscellaneous works: 1990-2002

Paradoxes without self-reference

Self-reference is apparently crucial in the whole matter of paradoxes and fixed point theorems represent the mathematically literate way of rephrasing most paradoxical arguments. So it may be of interest to inquire to what extent self-reference is essential or it can be eliminated in favor of alternatives. Historically, as we already know from §2.5 and §5.6., there are genuine paradoxes arising from unfoundedness or ungroundedness. For instance, it has been remarked that using infinite levels, one can also obtain a semantical contradiction without self-reference. This is Yablo's paradox [Yablo, 1985], [Yablo, 1993].

Assume that there are infinitely many people a_0, a_1, a_2, ... and each one says the same sentence: "at least one of the people following me is lying". Then there is no classical assignment to the statements, since, if a sentence a_i is true, it must be followed by a false one, say a_k, for some $k > i$, which in turn must be followed *only by true statements*. Then a_{k+1} is true and hence followed by a false sentence: contradiction. Hence a_i must be false; so it is followed *only by true statements*; a fortiori, a_{i+1} is true, hence followed by a false sentence; contradiction.[242]

[241] To this aim, let us also mention: (i) contributions on circularity, truth and paradoxes, due to Gupta, Chapuis, and others, with developments of Gaifman's theory of pointers, in the collective volume [Chapuis and Gupta, 2000]; (ii) McDonald [2000] which introduces a theory of truth, which is comparable in expressive power to Kripke's, and, in addition has a notion of *being meaningful*, that "can internally express the meaningfulness of any meaningless expression can itself be meaningful expressed"; (iii) Field [2003]; (iv) the discussion of the analogy between the logic of truth and the logic of rational choice in Gaifman [1999].

[242] At a closer look, this is actually not a paradox without self-reference; for instance, if we try to formalize the very definition of the sequence $(a_i)_{i\in\omega}$ in the language of Peano arithmetic with a truth predicate T, it is natural to proceed self-referentially, selecting with the second recursion theorem (for primitive recursive indexes PRI) a PRI-index fixed point a, such that

$$[a](n) = \overbrace{\ulcorner(\exists k)(k > n \wedge \neg T([a](k))\urcorner}^{\cdot}$$

In the same vein, it may be of interest to mention a game-theoretic version of Mirimanoff's paradox: the so called *hypergame paradox* of Zwicker (see [Zwicker, 1987]). A game G between two players A and B is grounded if every game of G terminates after finitely many moves. The hypergame is: A chooses a grounded game G, B chooses a move in G and the game then follows the rules of G. G being grounded, also the hypergame is grounded. So A can choose the hypergame itself, and then B has to choose. But B can choose the hypergame again, and so on; so we get an infinite game carried out according to the rule of a grounded game ...

The general proof pattern behind the hypergame argument has been exploited by [Bernardi and D'Agostino, 1997] as a sort of general proof schema for proving results in recursion theory, set theory and the theory of Magari algebras.

A mathematical approach to the general issue *self-reference vs. unfoundedness* has been recently developed by Bernardi [2001]. He observes that, besides Yablo's and Hypergame paradoxes, several classical results - existence of a non r.e. set, Cantor's theorems on the non-denumerability of the reals and the non-existence of a surjection from a set onto its power set - can be transformed into unfoundedness arguments.

This leads to the natural idea of regarding *unfounded chains as generalized fixed points* and to investigate corresponding existence theorems (in analogy with standard fixed point theorems).

To make this precise, let $f : X \to X$: a sequence $(x)_{n\in\omega}$ of elements of X is an *unfounded chain* if $f(x_{n+1}) = x_n$ for every $n \in \omega$. The problem is to find conditions on X and f, guaranteeing the existence of an unfounded chain. Here is an example of a topological solution:

Theorem ([Bernardi, 2001, p. 167]). Let X be a compact Hausdorff space and let $f : X \to X$ be a continuous function. Then f has an unfounded chain.

On the proper logical side, there is an interesting application to Magari algebras:

Theorem ([Bernardi, 2001, p. 172]). Every term in a Magari algebra has an unfounded chain.

This is a generalization of the well-known fixed point theorem of provability logic, where $f(x) = x$ has solutions, provided every occurrence of x is boxed or positive in $f(x)$ (see [Boolos and Sambin, 1991], [Boolos, 1993]).

Category theoretic and type-theoretic paradoxes

In 1969 Lawvere [1969] observes that "the similarity between the famous arguments of Cantor, Russell, Gödel and Tarski (...) suggests that these arguments should all be special cases of a single theorem about a suitable kind of abstract structure". He goes on to offer a fixed point theorem in cartesian closed categories (in short c.c.c's).

This kind of conceptual frame has been later worked out with the categorical interpretation of untyped lambda calculus [243] and in a corresponding categorical treatment of fixed point constructions.

Recently, Pavlovic [1992] develops Lawvere's insight by introducing the notion of *paradoxical structure* over the type Ω of truth values in a cartesian closed category; again paradoxical structure are so devised that every morphism t from Ω to Ω in a given c.c.c $\mathbb{C}$ has a fixed point Y_t, in the sense that $t \circ Y_t = Y_t$.

[Pavlovic, 1992] also introduces a notion of *generalized paradoxical structure*, in order to cover recent paradoxes arising from (polymorphic) type theories, in particular Reynolds's construction in [Reynolds, 1984] of a type $\mathcal{A} \simeq \mathcal{PPA}$.[244]

In type-theoretic context, there is a well-known paradox, due to Girard, for Martin-Löf's (unpublished version of) type theory with the type of all types [Troelstra and van Dalen, 1988] ch.11, 7.4. If the type U of all types exists, it is possible to define the type of all well-founded relations on elements of U and this yields a contradiction, which is a type-theoretic version of the paradoxes of Burali-Forti and Mirimanoff.

Since the mid 80's, Girard-type paradoxes have become important as consistency tests of certain impredicative extensions of polymorphic lambda calculus and its generalization (the theory of constructions). For instance, Coquand [1986] has proved that the theory of constructions with four levels as well as its extension with the Σ constructor are inconsistent; see also [Coquand, 1994]. A general point is that we cannot accept simultaneously second order impredicative quantification as granted, say, by the axiom stating that the collection of propositions forms a type and interpret existential quantification as generalized disjoint sum (with projections; this the so-called strong Σ). For additional discussion of the problem, see [Jacobs, 1989].

It is worth mentioning that all these paradoxes could also be naturally grouped with those of §7.1.

Berry's paradox and incompleteness

In 1966 P. Vopenka [1966] gave an alternative proof of the second incompleteness theorem for the Bernays-Gödel axiomatic theory by exploiting Berry's paradox (§2.3). Further work can be found in the textbook on set theory of Krivine [1972].

More recently, an elegant application of Berry's paradox is to be found in Boolos [1989], where incompleteness is proved in the form: *there is no algorithm whose output contains all true statements of arithmetic and no false ones.* The idea is to render the notion of nameability, which is crucial for Berry's argument, in the

[243] See Lambek's work, [Lambek and Scott, 1986] with the notion of C-monoid, and Longo and Asperti [1991].

[244] More generally, it has been established that, for each inhabited type B in polymorphic lambda calculus with equality, there exists a type A such that

$$(A \to B) \to B \simeq A$$

arithmetical language and this can be achieved using some precise version of the notion of algorithm (Turing machines, etc.). So, assuming that M is a machine whose output contains no false statement, one says that $F(x)$ names n, provided M outputs the (code of the formula) $(\forall x(x = \bar{n} \leftrightarrow F(x))$. This implies a smooth and elegant rephrasing of Berry's argument; remarkably, Boolos' argument does not use diagonalization.

Boolos's version can be criticized, since it is weaker than the standard one and hence it does not yield the second incompleteness theorem. Kikuchi [1994] shows that the difficulty can be overcome, if one understands the relation "$F(x)$ *names* n" as:

$$PA \vdash F(\bar{n}) \wedge (\forall x)(F(x) \rightarrow x = \bar{n})$$

The second incompleteness theorem is then proved by model-theoretic methods, in particular using the existence of *definable end extensions* for models of $PA + Cons(T)$, where T is a recursively axiomatizable extension of PA, $Cons(T)$ is the formalized consistency statement for T. This is a consequence of the so-called *arithmetized completeness theorem* and takes inspiration from Kreisel's model theoretic proof of the second incompleteness theorem.

Usually, these model theoretic proofs need stronger metamathematical assumptions, but a careful analysis, as carried out in [Kikuchi and Tanaka, 1994], shows that the relevant work can be done in the well-known fragment of second order arithmetic, WKL_0, which is conservative over primitive recursive arithmetic by a theorem of Friedman (see [Simpson, 1999]).

The Berry paradox is linked to the incompleteness phenomena also because of work (going back to the sixties and the seventies) in the so-called Kolmogorov complexity and algorithmic information theory (due to Kolmogorov, Solomonoff, Chaitin; see for references [Chaitin, 1995], [Odifreddi, 1989]).

Indeed, let the Kolmogorov complexity $K(n)$ of a number n be defined as the least (index of a) Turing machine e such that $\{e\}(0) \simeq n$ and call a number n *random* if and only if $n \leq K(n)$.

Kolmorogov proved in 1963 that the set of random numbers is *simple* (so it is recursively enumerable, but its complement is infinite and does not contain infinite r.e. subsets). Remarkably, the proof is again a form of Berry's argument and Chaitin has shown how to exploit randomness to make the limitations of formal systems precise.

Kikuchi [1997] proved the second incompleteness theorem via Kolmogorov complexity. The formalized first incompleteness theorem takes the form: for every recursively axiomatizable extension T of PA, there exists a numeral e such that, provably in T,

$$Cons(T) \rightarrow (\forall x)(\neg Pr_T(\overbrace{\ulcorner e < K(x) \urcorner}^{\cdot})) \tag{41}$$

$$\omega - Cons(T) \rightarrow (\forall x)(e < K(x) \rightarrow \neg Pr_T(\overbrace{\ulcorner K(x) \leq e \urcorner}^{\cdot})) \tag{42}$$

(ω-Cons(T) formalizes ω-consistency). The unprovability of $Cons(T)$ is then implied by (41)-(42) with the arithmetized completeness theorem and the same techniques of [Kikuchi and Tanaka, 1994].

For an updated survey of new proofs of the incompleteness, found in 1990s, see [Kotlarski, 2003].

Neo-Fregean theories

Since the eighties, there has been an attempt to isolate those parts of the system of the Fregean *Grundgesetze der Arithmetik*, that can be vindicated as consistent and yet have enough strength to define a model of natural numbers. This has the consequence of inducing a critical reappraisal of logicistic foundations of arithmetic, where *numbers are viewed as extensions of concepts.*

Roughly, the prototype Fregean system can be regarded as an extension of second order logic with a function symbol designating an injective application ϵ from concepts into individuals, i.e. satisfying

(43) $\epsilon(F) = \epsilon(G) \leftrightarrow F \equiv G$

Here F, G, etc. stand for concept variables , while $\equiv$ means that $(\forall x)(Fx \leftrightarrow Gx)$. (43) corresponds to axiom V of Frege's *Grundgesetze der Arithmetik*, which allows a definition of membership as

$$x \in y \Leftrightarrow (\exists F)(y = \epsilon(F) \wedge Fx)$$

and the provability of the naive comprehension by means of full second order comprehension CA_2,[245] whence Russell's contradiction. In the literature one can find two strategies for repairing the system:

(i) restricting (43);

(ii) restricting CA_2.

Ad [(i)]. This direction was inaugurated in 1983 by C. Wright, followed then by Burgess, Hodes, Boolos, Heck and others. An idea of technical results and of the still ongoing research can be extracted from Boolos's work [1987] and from the essays in part II of [Boolos, 1998]. Roughly, the technical achievement can be summarized as follows. Let us consider the extension FA of second order logic with a binary relation η which makes sense of concepts (represented as second order variables) and individuals. Let $FeqG$ be the second order formula expressing the relation " there is a bijection between the objects falling under F and those falling under G". Add the axiom *Numbers*

$$(\forall F)(\exists! x)(\forall G)(G\eta x \leftrightarrow FeqG)$$

[245] I.e. the schema $(\exists F)(\forall x)(Fx \leftrightarrow \varphi)$, where φ is an arbitrary formula and F does not occur in φ.

Then the main theorem is that *FA is (finitistically) interpretable in second order arithmetic PA_2, and conversely FA yields an inner model of PA_2*. Thus FA is powerful enough to recover the crucial parts of Frege's construction. In particular, *Numbers* implies the so-called Hume Principle, that is, there exists an operation $F \mapsto NF$ satisfying

$$NF = NG \leftrightarrow FeqG$$

Ad [(ii)] The instance of CA_2 which is responsible of contradictions, is Σ^1_1 (see the definition of $\in$). There are results ensuring the consistency of the Fregean subsystems based on comprehension restricted to predicative formulas (Heck [1996]) and to Δ^1_1-formulas (Ferreira and Wehmaier [2002]; see also Bell [1994]).

Quinean set theories

We already mentioned Quine's theories in §3.4 . Let us add that in the early fifties Quine's theory was used also as a basic logical system for mathematical logic and the foundations of mathematics. Later most work has dealt with the consistency problem. There are important results, due to Specker, Jensen, Grišin, Boffa, Crabbé, Forster and others. Here is a sample:

1. Specker [1962] establishes the equivalence of NF with standard type theory with extensionality plus *the schema of typical ambiguity*, which roughly states that two arbitrary types i and $i+1$ are isomorphic; Specker's construction leads to a structural counterpart of NF-models, reducing the consistency problem to the existence of models for type theory with *shifting automorphisms* (i.e. with automorphisms granting that typical ambiguity holds);

2. R. B. Jensen [1969] has shown the existence of models with shifting automorphisms, *provided extensionality is weakened to allow atoms*. This yields the consistency of a theory NFU and variants thereof with axioms of infinity and choice; [246] the proof requires models with indiscernibles and Ramsey's theorem;

3. Grišin proved that the full strength of NF is already given by NF_4, i.e. the system using the types 0, 1, 2, and 3; on the other hand, it is known that NF_3 is consistent and weak, see [Boffa, 1977];

4. Crabbé [1982] has proven the consistency of the impredicative subsystems NFI of NF, where a set $\{x \mid \varphi\}$ exists provided φ can be stratified using types not exceeding the type of the set itself.

Feferman [1974b] experiments with NF, in order to delineate a foundation for a general theory of structures and category theory. A systematic survey of NF is given by Forster [1995], while a reappraisal of the restricted system NFU with

[246] NF is inconsistent with choice and proves an axiom of infinity as known by Specker NFU is consistent relative to Peano arithmetic.

urelements is due to R. Holmes in the monograph [1998], where also a form of *stratified combinatory logic version* of NFU is investigated.

At last, NF has ramifications into the theory of concepts: there are contributions by Cocchiarella [1986], [1992] to property theory and to predication [1989], basically on extensional grounds and closer in spirit to the original Fregean ideas. This work leads to theories of concepts equiconsistent with Quinean set theories.

Recently, [Cantini, 2004] has investigated a Kripke-like theory of truth in a consistent, though fully impredicative fragment of Quine's set theory. In the new model, the set of all propositions and the set of all truths do exist, and, to a certain extent, the notion of truth has rather strong closure properties (this is motivated by the Russell paradox of §2.1; for alternative analyses, [Oksanen, 1999], [Cocchiarella, 2000]).

Positive set theories and hyperuniverses

We already know from §4.4 that non-uniform positive comprehension is consistent with extensionality. But it is important to observe that principles like GPC and its models occur in apparently unrelated fields, like the foundations of set theory and the mathematical theory of concurrency (for references regarding the applications, [Forti and Honsell, 1996b]).

New results strengthening [Forti and Hinnion, 1989] appear in [Forti and Honsell, 1996b] and in [Forti and Honsell, 1996a], which is mainly motivated by the consistency problems of GPC with choice principles.

The notion of *hyperuniverse* is introduced, in order to provide abstract structural counterparts of models by the comprehension principle GPC of [Forti and Hinnion, 1989] for the class GPF of *generalized positive formulas*.[247]

Hyperuniverses can be defined via algebraic closure conditions: a transitive set N is a hyperuniverse if and only if

1. N is closed under binary union and arbitrary intersections;
2. if $x \in N$, then $\{x\}, \bigcup x, \mathcal{P}(x) = \{u \in N \mid u \subseteq x\}, \mathcal{D}(x) = \{u \in N \mid u \cap x = \emptyset\}$ are all in N;
3. separation: if x and y are disjoint elements of N, there exists $z \in N$ such that $x \subseteq z$ and $y \subseteq (N - z)$.

On the other hand, there are *topological characterizations of hyperuniverses*: they can be naturally assigned a k-compact uniform k-topology, for some weakly compact strongly inaccessible cardinal, and they coincide with their exponential space. Furthermore, it is possible to give a general concrete construction of hyperuniverses (within models of set theory with choice and Anti-foundation AFA), which specializes at once to all the previously known hyperuniverses, like the topological models M_k of Malitz and the models N_k studied in [Forti and Honsell, 1996a].

[247] Recall that GPF is the least collection which contains prime formulas of the form $x = y$, $x \in y$ and is closed under disjunction, conjunction and quantification (both existential and bounded universal quantification).

GPC-theories can have great strength and their relations with choice is rather delicate; for instance, Esser [1999] shows that the Morse-Kelley impredicative class theory and GPC with the infinity axiom and a suitable closure schema due to Boffa are mutually interpretable. On the other hand, the same theory is inconsistent with the axiom of choice.

Topological models are also studied by Hinnion [1994], who tackles the consistency problem of naive comprehension with extensionality in partial logic and its dual (called here paradoxical; see §5.3).

Recently, Hinnion and Libert [2003] show that if uniform positive comprehension is restricted to =-free conditions, then it is consistent to add extensionality in the form: if two sets have the same elements, then they belong to the same sets (= is defined à la Leibniz). So all the inconsistency results involving uniform comprehension and extensionality (think of Gilmore's and Gordeev's proofs) *essentially depend on the fact that = is assumed as a primitive notion.*

Possible world semantics for modal notions conceived as predicates

This is a most recent work done by Halbach, Leitgeb and Welch [2003]. In spite of the well known negative results of [Montague, 1963], [Kaplan and Montague, 1960] (see section 6.1), there are good reasons to interpret modal notions with predicates:[248] in particular, the expressive power of the language is severely restricted when $\Box$ is a modal operator. It is then shown how to develop a Kripke semantics for the modality $\Box$ understood as a predicate (and hence $\Box A$ is properly $\Box(\ulcorner A \urcorner)$). A frame $\langle Wd, R\rangle$ consists of a set Wd of worlds and a binary relation R on Wd, where each world is a standard model of Peano Arithmetic, expanded with a denotation for $\Box$. Then $\Box A$ holds in a world i if and only if $k \models A$ in all models k, which are R-accessible from i.

As expected, paradoxes or limitative theorems (like the Liar, Gödel's second incompleteness theorem, Montague [1963] and McGee [1985]) transform into results, implying the non-admissibility of many frames, e.g., reflexive frames, for the predicate interpretation. Thus the main problem is to find sufficient and necessary conditions for the existence of a suitable interpretation of $\Box$ at any world (Characterization Problem CP).

Besides many non-trivial results and methods, already surveyed in §5 above, the (partial) solution of CP requires hard work and results of admissible set theory, higher recursion theory, ω-logic. Once more, it witnesses the technical sophistication of to-day formal semantics.

Logics with generalized quantifiers and paradoxes

This work is essentially due to Hintikka and his school ([Hintikka, 1995a], [Hintikka, 1998], [Sandu, 1998]). The starting point is that the Σ^1_1-fragment of the second order language of arithmetic can be characterized as an *independence friendly*

[248] Earlier work in this direction is due to Asher and Kamp [1989] and to Gupta and Belnap [1993].

(IF) first order language, "which permits that an existential quantifier ($\exists y$) is independent on a universal quantifier ($\forall x$) within whose scope it occurs. This IF-quantification is indicated with ($\exists y/\forall x$) and it allows a game theoretical semantics, where truth corresponds to the existence of suitable Skolem functions (or strategies); in particular ($\exists y/\forall x$) corresponds to the choice of a move disregarding any information about the move of the player associated with the universal quantification on x.

It is shown that the truth predicate for the IF-language is again expressible as an IF-formula and hence one obtains a semantically closed language in complexity Σ^1_1. The Liar paradox implies that in the IF-language the law of excluded middle fails, since there are games that are not determinate. Another feature of truth for the IF-language is that the definition is not compositional: the truth of a complex sentence is not fully determined by its simpler components.

Recent developments with contraction free logics

White [1987], [1993] has investigated a strengthening of Grišin's logic. He observes that a fixed point property holds in the system and proves inconsistency with extensionality, again in a natural deduction framework. In [Restall, 1993] there is an attempt to give an abstract characterization of those logics which support naive comprehension. Akama [1996] introduces a natural deduction calculus based on contraction free logic and untyped comprehension, which also includes a constructive negation in the sense of Nelson. The main result is a normal form theorem, which implies the consistency of the corresponding naive set theory (in a sense, this is a natural deduction counterpart [249] of Grišin's set theory). Petersen [2000] considers a contraction free calculus based on inclusion and abstraction as primitive notions, where logical notions can be defined. The calculus actually amounts to the intuitionistic fragment of Grišin (i.e. BCK-logic). After proving a cut elimination theorem and a fixed point theorem (compare with [Girard, 1998], [Cantini, 2003]), he faces the problem of founding natural numbers in a contraction free framework. There is a difficulty here, since proving the main closure conditions and induction for numbers does require some form of contraction. This leads to extend the basic calculus with suitable rules of induction, and the consistency proof becomes quite complex (it is carried out à la Gentzen using an ordinal assignment to derivations and transfinite induction up to ω^ω). Interestingly, the

[249] In this connection, let us recall also a much earlier suggestion in proof-theoretic vein, inspired by Fitch [1952], and given by Prawitz in his monograph [1965]. There he considers natural deduction with the additional abstraction rules:

$$(44)\quad \frac{A[x := t]}{t \in \{x \mid A\}} \qquad \frac{t \in \{x \mid A\}}{A[x := t]}$$

He then observes that paradoxes involve deductions without normal form, and so he proposes to distinguish between *quasi-deductions* (arbitrary usual natural deductions in the extended calculus) and *deductions in the proper sense*, that have to be *normal*. So the lesson from paradoxes is that there are quasi-deductions, which cannot be normalized. This idea of a proof *theoretic characterization of paradoxes* has received an interesting development in the papers of Tennant [1982], [1995].

calculus becomes inconsistent as soon as the Łukasiewicz logic $Ł_\aleph$ is assumed (in contrast with [White, 1979]).

Contraction free truth. Skolem's idea of applying analytical techniques to solve classical paradoxical situations has been recently revived by Hájek, Paris and Shepherdson [2000]: they prove that there exist models of the theory $\mathsf{PAŁTr}_2$, i.e. Peano Arithmetic with self-referential truth, endowed with (a suitable version of) Ł-logic[250] and the schema

$$T(\lceil\varphi\rceil) \leftrightarrow \varphi$$

φ being an *arbitrary sentence* of the appropriate extended arithmetical language. However there is no way to expand the standard model of Peano arithmetic to a model of $\mathsf{PAŁTr}_2$. So we cannot have a natural truth notion for arithmetic even in fuzzy logic.[251] It is also shown that it is not possible to find a model of PA with Ł-logic and (some) inductive axioms for truth, since the resulting theory is contradictory.

REMARK 7. Incidentally, observe that levels of implication (see §4.3, [Flagg and Myhill, 1987b]) are a way of blocking contraction for implication.

ZFC in a linear environment. Proof-theoretic methods have been successfully applied by Shirahata [1996] in order to show that the theory LZF with naive comprehension and linear logic is a conservative extension of ZF^-, i.e. ZF without foundation. The proof makes use of a cut elimination argument (based on a suitable triple induction), and an extension to the set-theoretic language of Girard's translation $A \mapsto A^+$ of classical logic into linear logic: the main result is that, if LZF proves A^+, then ZF^- proves A.

7.2 "From the Liar Paradox to Computer Science"

Self-referential theories of truth and abstraction, though originally suggested by the Foundations of Mathematics, are tentatively applied to problems, arising from Theoretical Linguistics, Computer Science and Artificial Intelligence.

Concerning the first direction, there are applications of type free systems based on Feferman and Aczel work to *the process of nominalization in natural languages*. By *nominalization* one roughly means the process of creating a noun phrase out of sentences, infinitives, etc.; e.g. this is apparent in passing from "*the old Galileo lived in Arcetri*" to: "*that the old Galileo lived in Arcetri is widely known*". This linguistic process has non-trivial logical aspects for two reasons.

On one side, there is an analogy with the abstraction process and with the problem of representing classes or concepts by means of individuals, which is basic in Frege's approach and in naive set theory, as well as in recent proposals (think of Jäger's systems cited in section 4.4).

On the other side, it is apparent that nominalization often generates *intensional* contexts and calls for a theory of properties or concepts; see [Chierchia and

[250] It is the so called RPŁ∀-logic, see [Hájek *et al.*, 2000].
[251] An ω-inconsistency is also stated in [Restall, 1992]

Turner, 1988], also the collective volume [Chierchia *et al.*, 1989], [Turner, 1983], [Turner, 1987], [Turner, 1997], [Kammeradine, 1992a], [Kammeradine, 1992b], [Orilia, 2000], [Hamm and van Lambalgen, 2001]. Technically, there is a variety of methods at work here, from domain theory, to fixed point theory and revision theory. Also, the work of Bealer [1994], [Bealer and Mönnich, 1989] on property theory is naturally subsumed under the applications to natural languages.

There is an additional issue which is of importance for Linguistics and has played a relevant role since the early investigations on paradoxes: *partiality*. For an updated analysis, the reader is referred to [Fenstad, 1997].

A less direct, but equally relevant fact concerns the use of partial models, three valued semantics, fixed point methods and related proof theoretic tools to problems pertaining to the semantics of Logic Programming. Here some pertinent paper: [Kunen, 1987], [Fitting, 1986], [Fitting, 1989], [Fitting, 1993]; [Jäger and Stärk, 1995]. As in the case of the relation between inductive definability and Kripkean truth languages, there is a two-sided link: for instance, [Fitting, 1997] lifts to the theory of truth the notion of *stable model* coming from the semantics of logic programming. The latest contribution in this direction is to our knowledge [Fitting, 2002].

Formal systems incorporating some form of self-referential truth or predication together with a ground applicative structure have been considered for application in Artificial Intelligence, especially to the issue of knowledge representation ([Perlis, 1985], [Perlis, 1988], [Perlis, 1989], [Perlis and Subrahmanjan, 1994], [Turner, 1990]) and to the theory of reasoning agents. The systems in question exploit supervaluation methods or semi-inductive methods (see [Davies, 1991]).

Logical frameworks for Computer Science, suggested by Frege structures (Aczel, Mendler, Carlisle [Aczel *et al.*, 1991]) have been also investigated and taken into account, as well as illative theories of relations, in connection with situation semantics, [Plotkin, 1990].

Similarly, theories of operations and classes of the sort Feferman investigated in the seventies and in the eighties (see §4.4) have inspired logical systems, to be used as foundations of computation and (functional) programming (e.g. see [Talcott, 1992], [Talcott and Mason, 2002]).

It is worth recalling that the circularity issue is typically borderline, and it bears at the same time upon linguistics, computer science and artificial intelligence. This is exemplified by the applications of the anti-foundation axiom and the corresponding mathematics of non-well founded phenomena, as developed by Barwise and Moss [1996]. As to the link with situation theory, we refer the reader to the chapter of [Seligman and Moss, 1997] in [Benthem and Meulen, 1997]

Interestingly enough, there has been an attempt [Moschovakis, 1990] to solve Frege's puzzle of the Morning and Evening Star within abstract recursion theory. In particular, Moschovakis [1990] works out a proposal of making the distinction between *sense* (Sinn) and *reference*(Bedeutung) precise. The natural informal idea is that senses correspond to algorithms, while denotations correspond to values (the sense of an expression, just as an algorithm, allows us to grasp the meaning

of an expression and identify its denotation, i.e the value).

This has been formally implemented in the definition of an extension LPCR (lower predicate calculus with reflection) of first order logic with a formula constructor WHERE for *directly introducing self-reference*.

Formulas of LPCR have a denotational semantics, i.e. they are interpreted as partial monotone functionals (using essentially a strong Kleene semantics) and the operator WHERE corresponds to the definitions by simultaneous recursion. Kripke's notion of self-referential truth is easily represented in LPCR.

Besides standard denotational semantics, LPCR possesses *an intensional semantics*: i.e. to each LPCR-expression one can associate a normal form (corresponding to the *intension*), so that intensional equality is decidable.

7.3 Appendix I: truth and logical complexity, a summary

Getting around the theorem on the undefinability of truth requires attention to the logical complexity involved in the chosen semantical schema. The moral is that under suitable weakening of logical complexity, there exist languages which can define their own truth. Here is a summary of what has been achieved:

1. positive existential languages admit of a positive existential truth predicate: this is essentially proved by Myhill [1950];

2. Σ_2-languages can express their own truth definition; this has been proved by Magari [1975];

3. there exist several natural languages of complexity Π^1_1, typically the combinatorial system of Fitch, Kripke's languages;

4. there exist truth predicates of complexity Σ^1_1 for languages that are semantically closed, typically (i) those obtained by considering the greatest fixed points of the monotone operators of the previous point; (ii) those based on the so-called IF-languages.

5. there exist self-referential truth predicates of complexity Π^1_2.

ACKNOWLEDGEMENTS

Research for this work has been supported at different stages by MURST and the University of Florence. It originated more than ten years ago (an extended abstract dates back to 1993) under the stimulus of Professor D. van Dalen, who kindly invited the author to contribute with a paper on paradoxes to a project for a History of Mathematical Logic. Since then, due to the slow pace the plan developed, the work has been updated several times and has grown up by successive extensions and revisions, till its merging in the present Handbook. We are grateful to the Editors for hospitality in the new enterprise. We wish to thank Reinhard Kahle and Akihiro Kanamori for useful comments and corrections.

BIBLIOGRAPHY

[Ackermann, 1941] W. Ackermann. System der typefreien Logik I. *Forschungen zur Logik und zur exakten Wissenschaften*, Heft 7, 1941.

[Ackermann, 1950] W. Ackermann. Widerspruchsfreier Aufbau der Logik. Typenfreies System ohne Tertium non datur. *Journal of Symbolic Logic*, 15:33–37, 1950.

[Ackermann, 1952] W. Ackermann. *Widerspruchsfreier Aufbau einer typenfreien Logik I. Mathematische Zeitschrift*, 55:364–384, 1952.

[Ackermann, 1958] W. Ackermann. *Ein typenfreies System der Logik mit ausreichender logischen Anwendungsfähigkeit. Archive for Mathematical Logic*, 4:3–26, 1958.

[Ackermann, 1961] W. Ackermann. Grundgedanken einer typenfreien Logik. In A. Fraenkel, editor, *Essays on the Foundations of Mathematics*, pages 143–155. Magnes Press, Jerusalem, 1961.

[Aczel and Feferman, 1980] P. Aczel and S. Feferman. Consistency of the unrestricted abstraction principle using an intensional equivalence operator. In Hindley and Seldin [1980], pages 67–98.

[Aczel *et al.*, 1991] P. Aczel, P. Carlisle, and N. Mendler. Two frameworks of theories and their implementation in Isabelle. In G.M. Plotkin, editor, *Logical Frameworks*, pages 5–39, Cambridge, 1991. Cambridge University press.

[Aczel, 1977] P. Aczel. The strength of Martin-Löf's intuitionistic type theory with one universe. In S. Mietissen and J. Vaänänen, editors, *Proceedings of the Symposium in Mathematical Logic in Oulu 1974 and Helsinky 1975*, volume 2, pages 1–32. Dept. of Philosophy, University of Helsinky, 1977.

[Aczel, 1980] P. Aczel. Frege structures and the notions of proposition, truth and set. In J. Barwise, H.J. Keisler, and K. Kunen, editors, *The Kleene Symposium*, pages 31–59. North Holland, Amsterdam, 1980.

[Aczel, 1988] P. Aczel. *Non-Well-Founded Sets*. CSLI, Stanford, 1988.

[Akama, 1996] S. Akama. Curry's paradox in contractionless constructive logic. *Journal of Philosophical Logic*, 25:135–150, 1996.

[Anellis, 1991] I. Anellis. The First Russell Paradox. In T. Drucker, editor, *Perspectives on the History of Mathematical Logic*, pages 33–46. Birkhäuser, 1991.

[Antonelli, 1994a] G. A. Antonelli. Non-well-founded sets via revision rules. *Journal of Philosophical Logic*, 23, 1994.

[Antonelli, 1994b] G.A. Antonelli. The complexity of revision. *Notre Dame Journal of Formal Logic*, 35:67–72, 1994.

[Antonelli, 1994c] G.A. Antonelli. A revision-theoretic analysis of the arithmetical hierarchy. *Notre Dame Journal of Formal Logic*, 35:204–208, 1994.

[Antonelli, 1998] G. A. Antonelli. The complexity of revision. revised version, pp. 1-3, 1998.

[Asher and Kamp, 1989] N. Asher and H. Kamp. Self-reference, attitude and paradox. In Chierchia et al. [1989], pages 85–158.

[Asperti and Longo, 1991] A. Asperti and G. Longo. *Categories, Types and Structures*. The MIT Press, Cambridge, Mass., 1991.

[Baldwin, 1905] J. Baldwin. *Dictionary of Philosophy and Psychology*. Clarendon, Oxford, 1905.

[Barwise and Etchemendy, 1987] J. Barwise and J. Etchemendy. *The Liar. An Essay in Truth and Circularity*. Oxford University Press, Oxford, 1987.

[Barwise and Moss, 1996] J. Barwise and L. Moss. *Vicious Circles. On the Mathematics of Non-Wellfounded Phenomena*. CSLI Publications, Stanford, 1996.

[Barwise, 1975] J. Barwise. *Admissible sets and structures*. Springer, Heidelberg, 1975.

[Bealer and Mönnich, 1989] G. Bealer and U. Mönnich. Property theories. In Gabbay and Guenthner [1989], pages 133–251.

[Bealer, 1994] G. Bealer. Property theory: the type-free approach vs. the Church approach. *Journal of Philosophical Logic*, 23:139–171, 1994.

[Beeson, 1985] M. Beeson. *Foundations of Constructive Mathematics*. Springer, Berlin-Heidelberg, 1985.

[Behmann, 1931] H. Behmann. *Zu den Widersprüchen der Logik und der Mengenlehre. Jahresbericht der Deutschen Mathematiker-Vereinigung*, 40:37–48, 1931.

[Behmann, 1959] H. Behmann. *Der Prädikatenkalkül mit limitierten Variablen: Grundlegung einer natürlichen exakten Logik. Journal of Symbolic Logic*, 24:112–140, 1959.
[Bell, 1994] J.L. Bell. Fregean extensions of first-order theory. *Mathematical Logic Quarterly*, 40:27–30, 1994.
[Belnap, 1982] N. Belnap. Gupta's rule of revision theory of truth. *Journal of Philosophical Logic*, 11:103–116, 1982.
[Benthem and Meulen, 1997] J. Van Benthem and A. Ter Meulen, editors. *Handbook of Logic and Language*. North Holland, Amsterdam, 1997.
[Bernardi and D'Agostino, 1997] C. Bernardi and G. D'Agostino. Translating the hypergame paradox: remarks on the set of founded elements of a relation. *Journal of Philosophical Logic*, 25:545–557, 1997.
[Bernardi, 2001] C. Bernardi. Fixed points and unfounded chains. *Annals of Pure and Applied Logic*, 109:163–178, 2001.
[Bernays, 1973] P. Bernays, editor. *L. Nelson: Gesammelte Schriften in neun Bändern*, volume II. Hamburg, 1973.
[Bernstein, 1905a] F. Bernstein. *Die Theorie der reellen Zahlen. Jahresbericht der Deutschen Mathematker-Vereinigung*, 14:447–449, 1905.
[Bernstein, 1905b] F. Bernstein. *Über die Reihe der transfiniten Ordnungszahlen. Mathematische Annalen*, 60:187–193, 1905.
[Bernstein, 1905c] F. Bernstein. Z*um Kontinuumproblem. Mathematische Annalen*, 60:463–464, 1905.
[Betti, 1999] A. Betti. The Porohy on Dnepr. Leswenieskian roots of Tarski's semantics. In T. Childers, editor, *The Logic Yearbook 1998*, pages 99–109, Filosofia, Prague, 1999.
[Bochvar, 1937] D.A. Bochvar. Ob odnom tréhznačnom isčislénii i égo priménénii k anlizu paradocsov klassičéskogo rasširénnogo funknional'nogo isčisléniá. *Mathematicheski Sbornik*, 4 (46):287–308, 1937. (English translation by M. Bergmann: 'On a three-valued logical calculus and its applications to the analysis of the paradoxes of the classical extended functional calculus' in: *History and Philosophy of Logic* vol.2:87-112, 1981).
[Boffa, 1977] M. Boffa. The consistency problem for *NF*. *The Journal of Symbolic Logic*, 42:215–220, 1977.
[Boolos and Sambin, 1991] G. Boolos and G. Sambin. Provability: the emergence of a mathematical modality. *Studia Logica*, 50:1–23, 1991.
[Boolos, 1987] G. Boolos. The consistency of Frege's foundations of arithmetic. In J. J. Thomson, editor, *On Being and Saying: Essays for Richard Cartwright*, pages 3–20, Cambridge, Mass. MIT Press, 1987. MIT Press.
[Boolos, 1989] G. Boolos. A new proof of the Gödel incompleteness theorem. *Notice of the American Mathematical Society*, 36:388–390, 1989.
[Boolos, 1993] G. Boolos. *The logic of provability*. Cambridge University Press, Cambridge, 1993.
[Boolos, 1998] G. Boolos. *Logic, Logic and Logic*. Harvard University Press, Cambridge, Mass., 1998. R. Jeffrey, ed.
[Borel *et al.*, 1905] E. Borel, R. Baire, J. Hadamard, and H. Lebesgue. Cinq lettres sur la théorie des ensembles. *Bulletin de la Sociéteé Mathématique de France*, 33:261–273, 1905.
[Borel, 1904] E. Borel. Quelques remarques sur les principes de la théorie des ensembles. *Mathematische Annalen*, 59:514–516, 1904.
[Borel, 1908] E. Borel. Les paradoxes de la théorie des ensembles. *Ann. École Nor. Sup.*, 25, 1908. also in [Borel, 1914].
[Borel, 1914] E. Borel. *Leçons sur la théorie des fonctions*. Gauthier-Villars, Paris, second edition, 1914.
[Brady, 1971] R.T. Brady. The consistency of the axioms of abstraction and extensionality in a three-valued logic. *Notre Dame Journal of Formal Logic*, 12:447–452, 1971.
[Brouwer, 1907] L.E.J. Brouwer. *Over die Grondslagen der Wiskunde*. Dissertation, Amsterdam, 1907. (English translation in: *Collected Works I*, North Holland, Amsterdam, 1975).
[Burali-Forti, 1897a] C. Burali-Forti. Sulle classi ben ordinate. *Rendiconti del Circolo Matematico di Palermo*, 11:260, 1897.
[Burali-Forti, 1897b] C. Burali-Forti. Una questione sui numeri transfiniti. *Rendiconti del Circolo Matematico di Palermo*, 11:154–164, 1897.
[Burge, 1979] T. Burge. Semantical paradox. *Journal of Philosophy*, 76:169–198, 1979.

[Burge, 1982] T. Burge. The liar paradox: tangles and chains. *Philosophical Studies*, 41:352–366, 1982.

[Burgess, 1986] J. Burgess. The truth is never simple. *Journal of Symbolic Logic*, 51:663–681, 1986.

[Burgess, 1988] J. Burgess. Addendum to: "The truth is never simple". *Journal of Symbolic Logic*, 53:390–392, 1988.

[Cain and Damnjanovic, 1991] J. Cain and Z. Damnjanovic. On the weak Kleene scheme in Kripke's theory of truth. *Journal of Symbolic Logic*, 56:1452–1468, 1991.

[Cantini and Minari, 1999] A. Cantini and P. Minari. Uniform inseparability in explicit mathematics. *The Journal of Symbolic Logic*, 64:313–326, 1999.

[Cantini, 1980] A. Cantini. A note on three-valued logic and Tarski theorem on truth definitions. *Studia Logica*, 39:405–415, 1980.

[Cantini, 1989] A. Cantini. Notes on formal theories of truth. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 35:97–130, 1989.

[Cantini, 1990] A. Cantini. A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic*, 55:244–259, 1990.

[Cantini, 1993] A. Cantini. Extending first order theory of combinators with self-referential truth. *Journal of Symbolic Logic*, 58:447–513, 1993.

[Cantini, 1995] A. Cantini. Levels of truth. *Notre Dame Journal of Formal Logic*, 36:185–213, 1995.

[Cantini, 1996] A. Cantini. *Logical Frameworks for Truth and Abstraction*. North Holland, Amsterdam, 1996.

[Cantini, 1997] A. Cantini. Proof-theoretic aspects of self-referential truth. In M. Dalla Chiara, K.Doets, and D.Mundici, editors, *Logic and Scientific Methods, I*, pages 7–27. Reidel, Dordrecht, 1997.

[Cantini, 2003] A. Cantini. The undecidability of Grišin's set theory. *Studia Logica*, 74:345–368, 2003.

[Cantini, 2004] A. Cantini. *On a Russellian paradox about propositions and truth*. In Link [2004], pages 259–284.

[Cantor, 1932;1962] G. Cantor. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Berlin; Hildesheim, 1932;1962.

[Cantor, 1991] G. Cantor. *Briefe*. Olms, Berlin–Hildesheim, 1991. edited by H. Meschkowski and W. Nilson.

[Carnap, 1932-33] R. Carnap. *Die logizistiche Grundlegung der Mathematik. Erkenntins*, 2:91–105, 1932-33.

[Carnap, 1934a] R. Carnap. *Die Antinomien und die Unvollständigkeit der Mathematik. Monatshefte für Mathematik und Physik*, 41:263–284, 1934.

[Carnap, 1934b] R. Carnap. *Die logische Syntax der Sprache*. Springer, Berlin, 1934.

[Chaitin, 1995] G. Chaitin. The Berry paradox. *Complexity*, 1:26–30, 1995.

[Chang, 1963] C.C. Chang. The axiom of comprehension in infinite-valued logic. *Mathematica Scandinava*, 13:19–39, 1963.

[Chang, 1965] C.C. Chang. Infinite-valued logic as a basis for set theory. In Y. Bar Hillel, editor, *Logic, Methodology and Philosophy of Science: Proceedings of the 1964 International Congress*, pages 93–100. North Holland, Amsterdam, 1965.

[Chapuis and Gupta, 2000] A. Chapuis and A. Gupta, editors. *Circularity, Definition and Truth*. Indian Council of Philosophical Reserach, New Dehli, 2000.

[Chapuis, 1996] A. Chapuis. Alternative revision theories of truth. *Journal of Philosophical Logic*, 24:399–423, 1996.

[Chierchia and Turner, 1988] G. Chierchia and R. Turner. Semantics and property theory. *Linguistics and Philosophy*, 11:261–302, 1988.

[Chierchia *et al.*, 1989] G. Chierchia, B. Partee, and R. Turner, editors. *Properties, Types and Meaning I*, Studies in Linguistics and Philosophy, Dordrecht, 1989. Reidel.

[Church, 1932] A. Church. A set of postulates for the foundations of logic (1st paper). *Annals of Mathematics*, 33:346–366, 1932.

[Church, 1933] A. Church. A set of postulates for the foundations of logic (2nd paper). *Annals of Mathematics*, 33:839–864, 1933.

[Church, 1934] A. Church. The Richard's paradox. *American Mathematical Monthly*, 41:356–361, 1934.

[Church, 1941] A. Church. *The Calculi of λ-Conversion*. Princeton University Press, Princeton, NJ, 1941.

[Church, 1976] A. Church. Comparison of Russell's resolution of the semantical antinomies. *Journal of Symbolic Logic*, 41:747–760, 1976.

[Chwistek, 1921] L. Chwistek. Antinomje logiki formalnej. *Przeglqd Filozoficzny*, 24:164–171, 1921. (English translation by Z. Jordan: Antinomies of formal logic, in: S. McCall, *Polish Logic 1920-1939*, Clarendon Press, Oxford, 1967, pg.338-345).

[Chwistek, 1922] L. Chwistek. Über die Antinomien der Prinzipien der Mathematik. *Mathemathische Zeitschrift*, 14:236–243, 1922.

[Chwistek, 1926] L. Chwistek. Über die Hypothesen der Mengenlehre. *Mathemathische Zeitschrift*, 25:439–473, 1926.

[Chwistek, 1929] L. Chwistek. *Neue Grundlagen der Logik und Mathematik*. *Mathemathische Zeitschrift*, 30:704–724, 1929.

[Chwistek, 1932] L. Chwistek. *Neue Grundlagen der Logik und Mathematik*, Zweite mitteilung. *Mathemathische Zeitschrift*, 34:527–534, 1932.

[Chwistek, 1933] L. Chwistek. *Die nominalistische Grundlegung der Mathematik*. *Erkenntnis*, 3:367–388, 1933.

[Chwistek, 1948] L. Chwistek. *The Limits of Science*. Kegan Paul, London, 1948. (English edition of *Granice Nauki*).

[Cocchiarella, 1986] N. Cocchiarella. *Logical Investigation of Predication Theory and the Problem of Universals*. Bibliopolis, Naples, 1986.

[Cocchiarella, 1989] N. Cocchiarella. Formal theories of predication. In Gabbay and Guenthner [1989], pages 254–326.

[Cocchiarella, 1992] N. Cocchiarella. Cantor's power-set theorem versus Frege's double-correlation thesis. *History and Philosophy of Logic*, 13:179–201, 1992.

[Cocchiarella, 2000] N. Cocchiarella. Russell's paradox of the totality of propositions. *Nordic Journal of Philosophical Logic*, 5:25–37, 2000.

[Coquand, 1986] T. Coquand. An analysis of Girard's paradox. In *LICS'86*, pages 227–236. IEEE Press, 1986.

[Coquand, 1994] T. Coquand. A new paradox in type theory. In D. Prawitz, B. Skyrms, and D. Westerstahl, editors, *LMPS'90*, pages 555–571. Elsevier Science, Amsterdam, 1994.

[Crabbé, 1982] M. Crabbé. *On the consistency of an impredicative subsystem of Quine's NF*. *Journal of Symbolic Logic*, 47:131–136, 1982.

[Curry, 1930] H.B. Curry. *Grundlagen der kombinatorischen Logik*. *American Journal of Mathematics*, 52:509–536, 739–834, 1930.

[Curry, 1934a] H.B. Curry. Functionality in combinatory logic. In *Proceedings of the National Academy of Sciences of the USA*, volume 20, pages 584–90, 1934.

[Curry, 1934b] H.B. Curry. Some properties of equality and implication in combinatory logic. *Annals of Mathematics*, 35:849–860, 1934.

[Curry, 1941] H.B. Curry. The paradox of Kleene and Rosser. *Transactions of the American Mathematical Society*, 41:454–516, 1941.

[Curry, 1942a] H.B. Curry. The combinatory foundations of mathematical logic. *Journal of Symbolic Logic*, 7:49–64, 1942.

[Curry, 1942b] H.B. Curry. The inconsistency of certain formal logics. *Journal of Symbolic Logic*, 7:115–117, 1942.

[Davies, 1991] N. Davies. A first-order theory of truth, knowledge and belief. In J. van Eijck, editor, *Logic in AI*, pages 170–179, Springer, Berlin, 1991.

[Dixon, 1907] A.C. Dixon. On well-order aggregates. *Proceedings of the London Mathematical Society (2nd series)*, 4:18–20, 1907.

[Drucker, 1991] T. Drucker, ed. *Perspectives on the History of Mathematical Logic*, Birkhäuser, 1991.

[Epstein, 1992] R. Epstein. A theory of truth based on a medieval solution of the liar paradox. *History and Philosophy of Logic*, 13:144–177, 1992.

[Esser, 1999] O. Esser. On the consistency of a positive theory. *Mathematical Logic Quarterly*, 45:105–116, 1999.

[Feferman, 1964] S. Feferman. Systems of predicative analysis I. *Journal of Symbolic Logic*, 29:1–30, 1964.

[Feferman, 1974a] S. Feferman. Non-extensional type free theories of partial operations and classifications I. In J. Diller and G.H. Müller, editors, *Proof Theory Symposium*, volume 500 of *Springer Lecture Notes in Mathematics*, pages 73–118, 1974.
[Feferman, 1974b] S. Feferman. Some formal systems for the unlimited theory of structures and categories. *The Journal of Symbolic Logic*, 39:374–375, 1974. abstract.
[Feferman, 1975] S. Feferman. A language and axioms for explicit mathematics. *Springer Lecture Notes in Mathematics*, 450:87–139, 1975.
[Feferman, 1977] S. Feferman. Categorical foundations and foundations for category theory. In R. Butts and J. Hintikka, editors, *Logic, Foundations of Mathematics and Computability*, pages 149–169, Reidel, Dordrecht, 1977.
[Feferman, 1979] S. Feferman. Constructive theories of functions and classes. In M. Boffa and D. van Dalen, editors, *Logic Colloquium '78*, pages 159–224. North Holland, Amsterdam, 1979.
[Feferman, 1982] S. Feferman. Iterated inductive fixed point theories: application to Hancock's conjecture. In G. Metakides, editor, *Patras Logic Symposion*, pages 171–195. North Holland, Amsterdam, 1982.
[Feferman, 1984] S. Feferman. Towards useful type-free theories I. *Journal of Symbolic Logic*, 49:75–111, 1984.
[Feferman, 1991] S. Feferman. Reflecting on incompleteness. *Journal of Symbolic Logic*, 56:1–49, 1991.
[Feferman, 2005] S. Feferman. Predicativity. In S. Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 590–625, Oxford, 2005. Oxford University Press.
[Fenstad, 1964] J.E. Fenstad. On the consistency of the axiom of comprehension in the Łukasiewicz infinite-valued logic. *Mathematica Scandinava*, 14:64–74, 1964.
[Fenstad, 1997] J.E. Fenstad. Partiality. In Benthem and Meulen [1997], pages 649–682.
[Ferreira and Wehmeier, 2002] F. Ferreira and K. Wehmeier. *On the consistency of the* Δ^1_1*-fragment of Frege's Grundgesetze. Journal of Philosophical Logic*, 31(2002):301–311, 2002.
[Field, 2003] H. Field. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic*, 32:139–177, 2003.
[Finsler, 1925] P. Finsler. *Gibt es Widersprüche in der Mathematik? Jahresbericht der Deutschen Mathematiker-Vereinigung*, 34:143–155, 1925.
[Finsler, 1926a] P. Finsler. *Formale Beweise und die Unentscheidbarkeit. Mathematische Zeitschrift*, 25:676–682, 1926.
[Finsler, 1926b] P. Finsler. *Über die Grundlagen der Mengenlehre. Mathematische Zeitschrift*, 25:683–713, 1926.
[Fitch, 1936] F.B. Fitch. A system of formal logic without an analogue to Curry W-operator. *Journal of Symbolic Logic*, 1:92–100, 1936.
[Fitch, 1942] F.B. Fitch. A basic logic. *Journal of Symbolic Logic*, 7:105–114, 1942.
[Fitch, 1944] F.B. Fitch. Representations of calculi. *Journal of Symbolic Logic*, 9:57–62, 1944.
[Fitch, 1946] F.B. Fitch. Self-reference in philosophy. *Mind*, 55:64–73, 1946.
[Fitch, 1948] F.B. Fitch. An extension of basic logic. *Journal of Symbolic Logic*, 13:95–106, 1948.
[Fitch, 1949] F.B. Fitch. The Heine-Borel theorem in extended basic logic. *Journal of Symbolic Logic*, 13:9–15, 1949.
[Fitch, 1952] F.B. Fitch. *Symbolic Logic*. Ronald Press, New York, 1952.
[Fitch, 1954a] F.B. Fitch. A definition of negation in extended basic logic. *Journal of Symbolic Logic*, 19, 1954.
[Fitch, 1954b] F.B. Fitch. A simplification of basic logic. *Journal of Symbolic Logic*, 18:317–325, 1954.
[Fitch, 1958] F.B. Fitch. Quasi-constructive foundations for mathematics. In A. Heyting, editor, *Constructivity in Mathematics*, pages 26–36. North Holland, Amsterdam, 1958.
[Fitch, 1963] F.B. Fitch. The system CΔ of combinatory logic. *Journal of Symbolic Logic*, 28:87–97, 1963.
[Fitch, 1964] F.B. Fitch. Universal metalanguages in philosophy. *Review of Metaphysics*, 17:396–402, 1964.
[Fitch, 1967a] F.B. Fitch. A complete and consistent modal set theory. *Journal of Symbolic Logic*, 32:93–103, 1967.
[Fitch, 1967b] F.B. Fitch. A theory of logical essences. *The Monist*, 51:104–109, 1967.

[Fitch, 1980] F.B. Fitch. A consistent combinatory logic with an inverse to equality. *Journal of Symbolic Logic*, 45:529–543, 1980.
[Fitting, 1986] M. Fitting. Notes on the mathematical aspects of Kripke's theory of truth. *Notre Dame Journal of Formal Logic*, 27:75–88, 1986.
[Fitting, 1989] M. Fitting. Bilattices and the theory of truth. *Journal of Philosophical Logic*, 18:225–256, 1989.
[Fitting, 1993] M. Fitting. The family of stable models. *Journal of Logic Programming*, 17:197–225, 1993.
[Fitting, 1997] M. Fitting. A theory of truth that prefers falsehood. *Journal of Philosophical Logic*, 26:477–500, 1997.
[Fitting, 2002] M. Fitting. Fixpoint semantics for logic programming: a survey. *Theoretical Computer Science*, 278:25–51, 2002.
[Flagg and Myhill, 1987a] R. Flagg and J. Myhill. An extension of Frege structures. In M. Dekker, editor, *Lecture Notes in Pure and Applied Mathematics*, volume 106, pages 197–217, New York, 1987.
[Flagg and Myhill, 1987b] R. Flagg and J. Myhill. Implications and analysis in classical Frege structures. *Annals of Pure and Applied Logic*, 34:33–85, 1987.
[Flagg and Myhill, 1989] R. Flagg and J. Myhill. A type-free system extending (ZFC). *Annals of Pure and Applied Logic*, 43:79–97, 1989.
[Flagg, 1987] R. Flagg. κ-continuos lattices and comprehension principles for Frege structures. *Annals of Pure and Applied Logic*, 36:1–16, 1987.
[Forster, 1995] T. E. Forster. *Set Theory with a Universal Set*. Clarendon Press, Oxford, 1995. 2nd edition.
[Forti and Hinnion, 1989] M. Forti and R. Hinnion. The consistency problem for positive comprehension principles. *Journal of Symbolic Logic*, 54:1401–1418, 1989.
[Forti and Honsell, 1983] M. Forti and F. Honsell. Set theory with free construction principles. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze, Serie IV*, 10:493–522, 1983.
[Forti and Honsell, 1996a] M. Forti and F. Honsell. Choice principles in hyperuniverses. *Annals of Pure and Applied Logic*, 77:35–52, 1996.
[Forti and Honsell, 1996b] M. Forti and F. Honsell. A general construction of hyperuniverses. *Theoretical Computer Science*, 156:203–215, 1996.
[Frege, 1962] G. Frege. *Grundgesetze der Arithmetik I, II*. Olms, Hildesheim, 1962.
[Frege, 1976] G. Frege. *Wissenschaftslicher Briefswechsel*. F. Meiner, Hamburg, 1976.
[Frege, 1984^2] G. Frege. *Nachgelassene Schriften*. F. Meiner, Hamburg, 1984^2.
[Friedman and Sheard, 1987] H. Friedman and M. Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21, 1987.
[Friedman and Sheard, 1991] H. Friedman and M. Sheard. The disjunction and existence properties for axiomatic systems of truth. *Annals of Pure and Applied Logic*, 56:1–49, 1991.
[Gabbay and Guenthner, 1989] D. Gabbay and F. Guenthner, editors. *Handbook of Philosophical Logic IV*. Reidel, Dordrecht, 1989.
[Gaifman, 1988] H. Gaifman. Pointers semantics: solution to self-referential puzzles I. In M. Vardi, editor, *TARK'88*, pages 43–59, 1988.
[Gaifman, 1992] H. Gaifman. Pointers to truth. *Journal of Philosophy*, 89:223–261, 1992.
[Gaifman, 1999] H. Gaifman. Self-reference and the acyclicity of rational choice. *Annals of Pure and Applied Logic*, 96:117–140, 1999.
[Gasser and Volken, 1975] J. Gasser and H. Volken, editors. *Logic and Set Theory in 20th century Switzerland*, Bern, 1975.
[Geach, 1955] P. Geach. On insolubilia. *Analysis*, 15:71–72, 1955.
[Gilmore, 1974] P. C. Gilmore. The consistency of partial set theory without extensionality. In T. Jech, editor, *Axiomatic set theory II*, volume 13 of *Proceedings of Symposia in Pure Mathematics*, pages 147–153. A.M.S., Providence R.I, 1974.
[Gilmore, 1980] P. C. Gilmore. Combining unrestricted abstraction with universal quantification. In Hindley and Seldin [1980], pages 99–123.
[Gilmore, 1986] P. C. Gilmore. Natural deduction based set theories: a new resolution of the old paradoxes. *Journal of Symbolic Logic*, 51:393–411, 1986.
[Gilmore, 2001] P. C. Gilmore. An intensional type theory: Motivation and cut- elimination. *Journal of Symbolic Logic*, 66:383–400, 2001.
[Girard, 1987a] J. Y. Girard. Linear logic. *Theoretical Computer Science*, 50:1–102, 1987.

[Girard, 1987b] J. Y. Girard. *Proof Theory and Logical Complexity*. Biliopolis, Napoli, 1987.
[Girard, 1998] J. Y. Girard. Light linear logic. *Information and Computation*, 143:175–204, 1998.
[Gödel, 1931] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.
[Gödel, 1944] K. Gödel. Russell's mathematical logic. In P. A. Schilpp, editor, *The Philosophy of Bertrand Russell*, pages 125–153. Northwestern University Press, Evanston, Ill., 1944.
[Grattan-Guinness, 1978] I. Grattan-Guinness. How Bertrand Russell discovered his paradox. *Historia Mathematica*, 5:127–137, 1978.
[Grelling and Nelson, 1908] K. Grelling and L. Nelson. *Bemerkungen zu den Paradoxien von Russell und Burali-Forti. Abhandlungen der Fries'chen Schule*, 2:301–334, 1908. (also in: L. Nelson, *Ges.Werke*).
[Griffin, 2004] N. Griffin. The Prehistory of Russell's Paradox. In Link [2004], pages 349–371.
[Grišin, 1974] V. N. Grišin. A non standard logic and its application to set theory (in Russian). In *Studies in formalized languages and non classical logics*, pages 135–171, Izdat, Nauka, Moskow, 1974.
[Grišin, 1981] V. N. Grišin. Predicate and set theoretic calculi based on logic without contraction rules (Russian). *Izvestiya Akademii Nauk SSSR Seriya Matematicheskaya*, 45(1):47–68, 1981. (English translation in: *Math. USSR Izv.*, vol. 18(1):41-59, 1982).
[Grue and Berline, 1997] K. Grue and C. Berline. A κ-denotational semantics for map theory in ZFC+SI. *Theoretical Computer Science*, 179:137–202, 1997.
[Gupta and Belnap, 1993] A. Gupta and N. Belnap. *The Revision Theory of Truth*. The MIT Press, Cambridge Mass., 1993.
[Gupta, 1982] A. Gupta. Truth and paradox. *Journal of Philosophical Logic*, 11:1–60, 1982.
[Hájek *et al.*, 2000] P. Hájek, J. Paris, and J. Shepherdson. The liar paradox and fuzzy logic. *Journal of Symbolic Logic*, 65:339–346, 2000.
[Halbach *et al.*, 2003] V. Halbach, H. Leitgeb, and P. Welch. Possible world semantics for modal notions conceived as predicates. *Journal of Philosophical Logic*, 32:179–223, 2003.
[Halbach, 1994] V. Halbach. A system of complete and consistent truth. *Notre Dame Journal of Formal Logic*, 35:311–327, 1994.
[Halbach, 1996] V. Halbach. *Axiomatische Wahrheitstheorien*. Akademie Verlag, Berlin, 1996.
[Halbach, 1999] V. Halbach. Conservative theories of truth. *Studia Logica*, 62:353–370, 1999.
[Halldén, 1949] S. Halldén. *The Logic of Nonsense*. Uppsala Universitets Årskrift, 1949.
[Hamm and van Lambalgen, 2001] F. Hamm and M. van Lambalgen. Event calculus, nominalisation, and the progressive. preprint, 1-77, 2001.
[Hayashi and Nakano, 1988] S. Hayashi and H. Nakano. *PX: a computational logic*. MIT Press, Cambridge, Mass., 1988.
[Heck, 1996] R. G. Heck, Jr. *The consistency of predicative fragments in Frege's Grundgesetze der Arithmetik. History and Philosophy of Logic*, 17:209–220, 1996.
[Herzberger, 1970] H. Herzberger. Paradoxes of grounding in semantics. *Journal of Philosophy*, 17:145–167, 1970.
[Herzberger, 1982a] H. Herzberger. Naive semantics and the liar paradox. *The Journal of Philosophy*, 79:479–497, 1982.
[Herzberger, 1982b] H. Herzberger. Notes on naive semantics. *Journal of Philosophical Logic*, 11:1–60, 1982.
[Hessenberg, 1906] G. Hessenberg. *Grundbegriffe der Mengenlehere*. Vandenhoek und Ruprecht, Göttingen, 1906. (also in: *Abhandlungen der Fries'chen Schule*, Neue Reihe, 1:479-706, 1906).
[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*. Springer, Berlin-Heidelberg, 1928.
[Hilbert and Bernays, 1939] D. Hilbert and P. Bernays. *Die Grundlagen der Mathematik I,II*. Springer, Berlin-Heidelberg, 1939.
[Hilbert, 1905] D. Hilbert. *Über die Grundlagen der Logik und der Arithmetik*. In *Verhandlungen des Dritten Internationalen-Mathematiker Kongresses*, pages 174–185. Teubner, Leipzig, 1905.
[Hilbert, 1918a] D. Hilbert. Axiomatisches Denken. *Mathematische Annalen*, 78:405–415, 1918.
[Hilbert, 1918b] D. Hilbert. Prinzipien der Mathematik. Vorlesung von D. Hilbert, Mathematisches Institut der Georg-August Universität Göttingen., 1918.
[Hindley and Seldin, 1980] R. Hindley and J. P. Seldin, editors. *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, New York, 1980. Academic Press.

[Hinnion and Libert, 2003] R. Hinnion and T. Libert. Positive abstraction and extensionality. *Journal of Symbolic Logic*, 68:828–836, 2003.
[Hinnion, 1994] R. Hinnion. Naive set theory with extensionality in partial logic and in paradoxical logic. *Notre Dame Journal of Formal Logic*, 35:15–40, 1994.
[Hintikka, 1995a] J. Hintikka. Definire la verità, tutta la verità, nient'altro che la verità. *Rivista di Filosofia*, 86:3–54, 1995.
[Hintikka, 1995b] J. Hintikka, editor. *From Dedekind to Gödel. Essays on the development of the foundation of mathematics*, Kluwer, Dordrecht, 1995.
[Hintikka, 1998] J. Hintikka. Truth definitions, Skolem functions and axiomatic set theory. *The Bulletin of Symbolic Logic*, 4:303–337, 1998.
[Hobson, 1905] E. W. Hobson. On the arithmetic continuum. *Proceedings of the London Mathematical Society*, 4:21–28, 1905.
[Holmes, 1998] M. R. Holmes. *Elementary Set Theory with a Universal Set*, volume 10 of *Cahiers du Centre de Logique*. Academia Bruylant, Louvain-La-Neuve, 1998.
[Howard, 1980] W. A. Howard. The formulas as types notion of construction. In Hindley and Seldin [1980], pages 480–490.
[Husserl, 1979] E. Husserl. *Aufsätze und Rezensionen (1890-1910)*. Nijhoff, The Hague, 1979.
[Jacobs, 1989] B. Jacobs. The inconsistency of higher order extensions of Martin-Löf's type theory. *Journal of Philosophical Logic*, 18:399–422, 1989.
[Jäger and Stärk, 1995] G. Jäger and R. Stärk. A proof theoretic framework for logic programming. In S.R. Buss, editor, *Handbook of Proof Theory*. Elsevier, 1995.
[Jäger *et al.*, 1999] G. Jäger, R. Kahle, A. Setzer, and T. Strahm. The proof-theoretic analysis of transfinitely iterated fixed point theories. *Journal of Symbolic Logic*, 64:53–67, 1999.
[Jäger *et al.*, 2001] G. Jäger, R. Kahle, and T. Strahm. Universes in explicit mathematics. *Annals of Pure and Applied Logic*, 109:141–162, 2001.
[Jäger, 1987] G. Jäger. Induction in the elementary theory of types and names. In E. Börger, H. Kleine-Büning, and M.M. Richter, editors, *Proceedings CSL'87*, volume 329 of *Springer Lecture Notes in Computer Science*, pages 118–128, 1987.
[Jäger, 1997] G. Jäger. Power types in explicit mathematics. *Journal of Symbolic Logic*, 62:1142–1146, 1997.
[Jech, 1973] T. Jech. *The Axiom of Choice*. Studies in Logic and the Foundations of Mathematics. North-Holland, Amsterdam, 1973.
[Jensen, 1969] R.B. Jensen. *On the consistency of a slight (?) modification of Quine's New Foundations*. In D. Davidson and J. Hintikka, editors, *Words and Objections: Essays on the work of W.V.O.Quine*, pages 278–291, Dordrecht, 1969. Reidel.
[Kahle, 2002] R. Kahle. *Universes over Frege Structures. Annals of Pure and Applied Logic*, pages 1–33, 2002.
[Kahle, 2004] R. Kahle. *David Hilbert über Paradoxien*. unpublished, 1-42, 2004.
[Kammeradine, 1992a] F. Kammeradine. Set theory and nominalisation I. *Journal of Logic and Computation*, 2:579–604, 1992.
[Kammeradine, 1992b] F. Kammeradine. Set theory and nominalisation II. *Journal of Logic and Computation*, 2:687–707, 1992.
[Kanamori, 2004] A. Kanamori. Zermelo and set theory. *The Bulletin of Symbolic Logic*, 10(4):487–553, 2004.
[Kaplan and Montague, 1960] D. Kaplan and R. Montague. A paradox regained. *Notre Dame Journal of Formal Logic*, 1:79–90, 1960.
[Kaye, 1991] R. Kaye. *Models of Peano Arithmetic*. Clarendon, Oxford, 1991.
[Kikuchi and Tanaka, 1994] M. Kikuchi and K. Tanaka. *On formalization of model-theoretic proofs of Gödel's theorems. Notre Dame Journal of Formal Logic*, 35:403–412, 1994.
[Kikuchi, 1994] M. Kikuchi. A note on Boolos' proof of the incompleteness theorem. *Mathematical Logic Quarterly*, 40:528–532, 1994.
[Kikuchi, 1997] M. Kikuchi. Kolmogorov complexity and the second incompleteness theorem. *Archive for Mathematical Logic*, 36:437–443, 1997.
[Kindt, 1978a] W. Kindt. The introduction of truth predicates into first order languages. In F. Guenthner and S.J. Schmidt, editors, *Formal Semantics and Pragmatics for Natural Languages*, pages 359–371, Reidel, Dordrecht, 1978.
[Kindt, 1978b] W. Kindt. *Über Sprachen mit Wahreitsprädikat*. In Ch. Habel and S. Kangiesser, editors, *Sprachdynamik und Sprachstruktur*, pages 205–238, Tübingen, 1978. Max Niemeyer.

[Kleene and Rosser, 1935] S.C. Kleene and J.B. Rosser. The inconsistency of certain formal logics. *Annals of Mathematics*, 36:630–637, 1935.
[Kleene, 1934] S.C. Kleene. Proofs by cases in formal logic. *Annals of Mathematics*, 35:529–544, 1934.
[Kleene, 1935] S.C. Kleene. A theory of positive integers in formal logic. *American Journal of Mathematics*, 57:153–173; 219–244, 1935.
[Kleene, 1952] S.C. Kleene. *Introduction to Mathematics*. Wolters-Nordhoff, 1952.
[Kleene, 1955] S.C. Kleene. On the forms of predicates in the theory of constructive ordinals (2nd paper). *American Journal of Mathematics*, 66:405–428, 1955.
[Komori, 1989] Y. Komori. Illative logic based on BCK-logic. *Mathematica Japanica*, 34:585–597, 1989.
[König, 1905a] J. König. *Über die Grundlagen der Mengenlehre und das Kontinuumsproblem. Mathematische Annalen*, 61:156–160, 1905.
[König, 1905b] J. König. *Zum Kontinuum-Problem. Jahresbericht der Deutschen Mathematiker-Vereinigung*, 14:177–180, 1905.
[König, 1906] J. König. *Über die Grundlagen der Mengenlehre und das Kontinuumsproblem. Zweite Mitteilung. Mathematische Annalen*, 63:217–221, 1906.
[Koons, 1992] R. Koons. *Paradoxes of Belief and Strategic Rationality*. Cambridge Studies in Probability, Induction and Decision theory. Cambridge University Press, Cambridge, 1992.
[Kotlarski, 2003] H. Kotlarski. The incompleteness theorems after 70 years. *Annals of Pure and Applied Logic*, 126:125–138, 2004.
[Kotlarski and Z.Ratajczyk, 1990] H. Kotlarski and Z.Ratajczyk. Inductive satisfaction classes. *Annals of Pure and Applied Logic*, 47:199–223, 1990.
[Kotlarski *et al.*, 1981] H. Kotlarski, S. Krajewski, and A. Lachlan. Construction of satisfaction classes for non-standard models. *Canadian Mathematical Bulletin*, 24 (3):283–293, 1981.
[Kotlarski, 1991] H. Kotlarski. Full satisfaction classes: a survey. *Notre Dame Journal of Formal Logic*, 32:573–579, 1991.
[Krajček, 1987a] J. Krajček. A possible modal formulation of comprehension scheme. *Zeitschrift für Mathematische ogik und Grundlagen der Mathematik*, 33:461–480, 1987.
[Krajček, 1987b] J. Krajček. Some results and problems in the modal set theory MST. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 33:123–134, 1987.
[Kreisel, 1960] G. Kreisel. La prédicativité. *Bulletin de la Societé Mathématique de France*, 88:371–391, 1960.
[Kremer, 1993] M. Kremer. The Gupta-Belnap systems $S^{\#}$ and S^{*} are not axiomatizable. *Journal of Philosophical Logic*, 34:583–596, 1993.
[Kripke, 1975] S. Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.
[Krivine, 1972] J.L. Krivine. *Théorie Axiomatique des Ensembles*. PUF, Paris, 1972.
[Kunen, 1987] K. Kunen. Negation in logic programming. *Journal of Logic Programming*, 4:231–245, 1987.
[Kuratowski and Tarski, 1931] K. Kuratowski and A. Tarski. Les opérations logiques et les ensembles projectifs. *Fundamenta Mathematicae*, 17:240–248, 1931.
[Lachlan, 1981] A. Lachlan. Full satisfaction classes and recursive saturation. *Canadian Mathematical Bulletin*, 24 (3):295–297, 1981.
[Lambek and Scott, 1986] J. Lambek and P. J. Scott. *Introduction to Higher Order Categorical Logic*. Cambridge University Press, Cambridge, 1986.
[Landini, 2004] G. Landini. Logicism's 'Insolubilia' and Their Solution by Russell's Substitutional Theory. In Link [2004], pages 373–399.
[Lawvere, 1969] F. W. Lawvere. Diagonal arguments and cartesian closed categories. In P. Hilton, editor, *Category Theory, Homology Theory and their Applications, II*, volume 92 of *Lecture Notes in Mathematics*, pages 134–145, Berlin-Heidelberg, 1969. Springer.
[Lemmon, 1963] E.J. Lemmon. A theory of attributes based on modal logic. *Acta Philosophica Fennica*, 16:95–122, 1963.
[Lesniewski, 1913] S. Lesniewski. Krytyca logicznej zasady wylaczonego środka. *Przegląd filozoficzny XVI*, pages 315–352, 1913. (English translation in: [Surma *et al.*, 1992], vol.1, pp. 47-85).
[Levi, 1902] B. Levi. Intorno alla teoria degli aggregati. *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere (seconda serie)*, 35:863–868, 1902.
[Levi, 1908] B. Levi. Antinomie logiche? *Annali di Matematica (terza serie)*, tomo 15:188–216, 1908.

[Lewis and Langford, 1932] C.I. Lewis and C.H. Langford. *Symbolic Logic*. The Century Co., New York, 1932. (Reprinted Dover, New York, 1952).
[Link, 2004] G. Link, editor. *One Hundred Years of Russell's Paradox*, Berlin, 2004. De Gruyter.
[Löb, 1955] M. Löb. Solution of a problem of Leon Henkin. *The Journal of Symbolic Logic*, 20:115–118, 1955.
[Lorenzen and Myhill, 1959] P. Lorenzen and J. Myhill. Constructive definitions of certain analytic sets of numbers. *Journal of Symbolic Logic*, 24:37–49, 1959.
[Löwe and Welch, 2000] B. Löwe and P. D. Welch. Set theoretic absoluteness and the revision theory of truth. Logic Colloquium, to appear, 2000.
[Maddy, 1983] P. Maddy. Proper classes. *Journal of Symbolic Logic*, 45:113–139, 1983.
[Magari, 1973] R. Magari. Meaning and truth in the Peano Arithmetic. *Atti dell'Accademia dei Lincei, Cl. Sci. Fis. Mat. Nat.*, 54:592–593, 1973.
[Magari, 1974] R. Magari. Su certe teorie non enumerabili. *Annali di Matematica Pura e Applicata*, 97:119–152, 1974.
[Magari, 1975] R. Magari. Significato e verità nell'aritmetica peaniana. *Annali di Matematica Pura e Applicata*, 103:343–368, 1975.
[Malitz, 1976] R.J. Malitz. *Set theory in which the axiom of foundation fails*. PhD thesis, University of California, Berekeley, California, 1976.
[Mancosu, 1999] P. Mancosu. Between Russell and Hilbert: Behmann on the foundations of mathematics. *The Bulletin of Symbolic Logic*, 5:302–330, 1999.
[Mancosu, 2003] P. Mancosu. The Russellian Influence on Hilbert and his School. *Synthèse*, 137:59–101, 2003.
[Markowski, 1976] G. Markowski. Chain-complete posets and directed sets with applications. *Algebra Universalis*, 6:53–68, 1976.
[Martin and Woodruff, 1975] R. L. Martin and P. Woodfruff. On representing 'True-in-L' in L. *Philosophia*, 5:217–221, 1975.
[Martin-Löf, 1975] P. Martin-Löf. Intuitionistic type theory: predicative part. In H. Rose and J. Shepherdson, editors, *Logic Colloquium '73*, pages 73–118, North Holland, Amsterdam, 1975.
[Martin, 1970] R. L. Martin, editor. *The Paradox of the Liar*. Yale University Press, Yale, 1970. (Reprinted by Ridgeview Press, Independence Ohio, 1978).
[Martin, 1970] R. L. Martin, editor. *Truth and the Liar Paradox*, Oxford Univeristy Press, 1984.
[McCarthy, 1988] T. McCarthy. Ungroundedness in classical languages. *Journal of Philosophical Logic*, 17:61–74, 1988.
[McDonald, 2000] B. E. McDonald. On meaningfulness and truth. *Journal of Philosophical Logic*, 29:1–50, 2000.
[McGee, 1985] V. McGee. How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14:399–410, 1985.
[McGee, 1991] V. McGee. *Truth, Vagueness and Paradox*. Hackett Publishing Company, Indianapolis-Cambridge, 1991.
[McGee, 1992] V. McGee. *Maximal consistent sets of instances of Tarski's Schema (T). Journal of Philosophical Logic*, 21(235-241), 1992.
[Minari, 1987] P. Minari. Some aspects of Kripke's theory of truth. Unpublished notes, 1987.
[Mirimanoff, 1917a] D. Mirimanoff. Les antinomies de Russell et de Burali-Forti et le probleme fondamentale de la théorie des ensembles. *L'enseignement mathématique*, 19:37–52, 1917.
[Mirimanoff, 1917b] D. Mirimanoff. Remarques sur la théorie des ensembles et les antinomies cantoriennes I. *L'enseignement mathématique*, 19:209–217, 1917.
[Mirimanoff, 1920] D. Mirimanoff. Remarques sur la théorie des ensembles et les antinomies cantoriennes II. *L'enseignement mathématique*, 21:29–52, 1920.
[Montague, 1963] R. Montague. Syntactical treatment of modality, with corollaries on reflection principles and finite axiomatizability. *Acta Philosophica Fennica*, 16:153–167, 1963.
[Moore and Garciadiego, 1981] G.H. Moore and A. Garciadiego. Burali-Forti's paradox: a reappraisal of its origins. *Historia Mathematica*, 8:319–350, 1981.
[Moore, 1995] G.H. Moore. The origin of Russell's paradox: Russell, Couturat and the antinomy of infinite number. In Hintikka [1995b], pages 215–239.
[Moschovakis, 1974] Y.N. Moschovakis. *Elementary Induction on Abstract Structures*. North Holland, Amsterdam, 1974.

[Moschovakis, 1990] Y.N. Moschovakis. Sense and reference as algorithm and value. In *Logic Colloquium' 90*, Springer Lecture Notes in Logic, pages 210–249, 1990.
[Mow-Shaw-Kwei, 1954] Mow-Shaw-Kwei. Logical paradoxes for many-valued systems. *Journal of Symbolic Logic*, 19:37–40, 1954.
[Myhill, 1950] J. Myhill. A system which can define its own truth. *Fundamenta Mathematica*, 37:190–192, 1950.
[Myhill, 1952] J. Myhill. A finitary metalanguage for extended basic logic. *Journal of Symbolic Logic*, 17:164–178, 1952.
[Myhill, 1975] J. Myhill. Levels of implications. In R. Barcan Marcus, R.L. Martin, and A. Ross, editors, *The Logical Enterprise*, pages 179–185, New Haven, 1975. Yale University Press.
[Myhill, 1984] J. Myhill. Paradoxes. *Synthèse*, 60:129–142, 1984.
[Negri, 1989] M. Negri. An autoapplicable truth predicate. *Bollettino dell'Unione Matematica Italiana*, 3-B:29–39, 1989.
[Nepeivoda, 1973] N.N. Nepeivoda. A new concept of predicative truth and definability. *Mathematical Notes*, 13:439–444, 1973. (English translation from: *Matematicheskie Zametki*, 13:735-745, 1973).
[Odifreddi, 1989] P. Odifreddi. *Classical Recursion Theory I.* Studies in Logic 125. North Holland, Amsterdam, 1989.
[Oksanen, 1999] M. Oksanen. The Russell-Kaplan paradox and other modal paradoxes: a new solution. *Nordic Journal of Philosophical Logic*, 4:73–93, 1999.
[Orilia, 2000] F. Orilia. Property theory and the revision theory of definitions. *The Journal of Symbolic Logic*, 65:212–245, 2000.
[Parsons, 1974] C. Parsons. The liar paradox. *Journal of Philosophical Logic*, 3:381–412, 1974. (Reprinted in: [Martin, 1970], pages 9-45, with a Postscript).
[Pavlovic, 1992] D. Pavlovic. On the structure of paradoxes. *Archive for Mathematical Logic*, 31:397–406, 1992.
[Peano, 1902-1906] G. Peano. Additione. *Revista de Matematica*, 8:143–157, 1902-1906.
[Peckhaus and Kahle, 2002] V. Peckhaus and R. Kahle. Hilbert's paradox. *Historia Mathematica*, 29:127–155, 2002.
[Peckhaus, 1990] V. Peckhaus. *Hilbert's Programm und Kritische Philosophie*. Vandenhoek and Ruprecht, Göttingen, 1990.
[Peckhaus, 2004] V. Peckhaus. Paradoxes in Göttingen. In Link [2004]. 501-515.
[Pelham and Urquhart, 1994] J. Pelham and A. Urquhart. Russellian propositions. In D. Prawitz, B. Skyrms, and D. Westerstahl, editors, *Logic, Methodology and Philosophy of Science IX*, pages 307–326, Amsterdam, 1994. Elsevier.
[Perlis and Subrahmanjan, 1994] D. Perlis and V.S. Subrahmanjan. Meta-languages, reflection principles and self-reference. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 2. Clarendon Press, Oxford, 1994.
[Perlis, 1985] D. Perlis. Language with self-reference I: foundations. *Artificial Intelligence*, 25:301–322, 1985.
[Perlis, 1988] D. Perlis. Language with self-reference II: knowledge, belief and modality. *Artificial Intelligence*, 34:179–212, 1988.
[Perlis, 1989] D. Perlis. Truth and meaning. *Artificial Intelligence*, 39:245–250, 1989.
[Petersen, 2000] U. Petersen. Logic without contraction as based on inclusion and unrestricted abstraction. *Studia Logica*, 64:3065–4003, 2000.
[Plotkin, 1990] G. Plotkin. An illative theory of relations. In R. Cooper, K. Mukai, and J. Perry, editors, *Situation Theory and its Applications*, pages 134–146, Stanford, 1990. CSLI.
[Plotkin, 1995] G. Plotkin. On a question of Friedman. pp. 1-16, 1995.
[Poincaré, 1905] H. Poincaré. Les mathématiques et la logique. *Revue de Métaphysique et Morale*, 13:815–835, 1905.
[Poincaré, 1906a] H. Poincaré. Les mathématiques et la logique. *Revue de Métaphysique et Morale*, 14:17–34, 1906.
[Poincaré, 1906b] H. Poincaré. Les mathématiques et la logique. *Revue de Métaphysique et Morale*, 14:294–317, 1906.
[Poincaré, 1909a] H. Poincaré. La logique de l'infini. *Revue de Métaphysique et Morale*, 17:461–482, 1909.
[Poincaré, 1909b] H. Poincaré. Réflexions sur les deux notes précédentes. *Acta Mathematica*, 32:195–200, 1909.

[Poincaré, 1910] H. Poincaré. Über transfiniten *Zahlen*. In *Sechs Vorträge über ausgewählte Gegenstände der reinen Mathematik und Physik*, pages 43–48, Leipzig, Berlin, 1910.
[Poincaré, 1912] H. Poincaré. La logique de l'infini. *Scientia*, 12:1–11, 1912.
[Prawitz, 1965] D. Prawitz. *Natural Deduction. A Proof-theoretical Study*. Almqvist and Wiksell, Stockholm, 1965.
[Priest, 1989] G. Priest. Reasoning about truth. *Artificial Intelligence*, 39:231–244, 1989.
[Quine, 1937] W.V.O. Quine. New foundations for mathematical logic. *American Mathematical Monthly*, 44:70–80, 1937.
[Ramsey, 1926] F. P. Ramsey. The foundations of mathematics. *Proceedings of the London Mathematical Society*, 25:338–384, 1926.
[Rathjen, 1998] M. Rathjen. Explicit mathematics with the monotone fixed point principle. *Journal fo Symbolic Logic*, 63:517–550, 1998.
[Rathjen, 1999] M. Rathjen. Explicit mathematics with the monotone fixed point principle ii: Models. *Journal fo Symbolic Logic*, 64:509–542, 1999.
[Reinhardt, 1985] W. Reinhardt. Remarks on significance and meaningful applicability. In L. P. de Alcantara, editor, *Mathematical Logic and Formal Systems*, volume 94 of *Lecture Notes in Pure and Applied Mathematics*, pages 227–242. Marcel Dekker, New York, 1985.
[Reinhardt, 1986] W. Reinhardt. Some remarks on extending and interpreting theories with a partial predicate for truth. *Journal of Philosophical Logic*, 15:219–251, 1986.
[Restall, 1992] G. Restall. Arithmetic and truth in Łukasiewicz's infinitely valued logic. *Logique et Analyse*, 139-140:303–313, 1992.
[Restall, 1993] G. Restall. How to be really contraction free. *Studia Logica*, 52:381–391, 1993.
[Reynolds, 1984] J. C. Reynolds. Polymorphysm is not set-theoretic. In *Semantics of Data Types*, Springer Lecture Notes in Computer Science 174, pp. 145–156, 1984.
[Richard, 1905] J. Richard. Les principes des mathématique et le problèmes des ensembles. *Revue Générale des Sciences Pures et Appliquées*, 16:541, 1905. (also in: *Acta Mathematica*, 30:295-296, 1906).
[Robinson, 1963] A. Robinson. On languages based on non-standard arithmetic. *Nagoya Mathematical Journal*, 22:83–107, 1963.
[Rogers, 1967] H. Rogers. *Theory of Recursive Functions and Effective Computability*. Mc Graw-Hill, New York, 1967.
[Rosser, 1939] J.B. Rosser. An informal exposition of proofs of Gödel's and Church's theorem. *Journal of Symbolic Logic*, 4:53–60, 1939.
[Rouilhan, 2004] P. De Rouilhan. Substitution and Types: Russell's Intermediate Theory. In Link [2004], pages 401–415.
[Russell, 1903] B. Russell. *The Principles of Mathematics*. Routledge and Kegan, London, 1903.
[Russell, 1905a] B. Russell. On denoting. *Mind*, 14:479–493, 1905.
[Russell, 1905b] B. Russell. On the substitutional theory of classes and relations. Received by the London Mathematical Society on 24 April 1906, (in: B. Russell, *Essays in Analysis*, edited by D. Lackey, Allen and Unwin, London, 1973), 1905.
[Russell, 1906a] B. Russell. Les paradoxes de la logique. *Revue de Mètaphysique et Morale*, 14:627–650, 1906.
[Russell, 1906b] B. Russell. On some difficulties in the theory of trasfinite numbers and order types. *Proceedings of the London Mathematical Society (2nd series)*, 4:29–53, 1906.
[Russell, 1908] B. Russell. Mathematical logic as based on the theory of types. *American Journal of Mathematics*, 30:222–262, 1908.
[Russell, 1993] B. Russell. *The Collected Papers of Bertrand Russell. Toward the Principles of Mathematics: 1900-02*. Routledge, London, 1993.
[Sainsbury, 1988] R.M. Sainsbury. *Paradoxes*. Cambridge University Press, Cambridge, 1988.
[Sandu, 1998] G. Sandu. IF- logic and truth definitions. *Journal of Philosophical Logic*, 27:143–164, 1998.
[Schönfinkel, 1924] M. Schönfinkel. *Über die Bausteine der mathematischen Logik. Mathematische Annalen*, 92:305–316, 1924.
[Schönflies, 1906] A. Schönflies. *Über die logischen Paradoxien der Mengenlere. Jahresbericht der Deutschen Mathematiker-Vereinigung*, 15:19–25, 1906.
[Schönflies, 1909] A. Schönflies. *Über eine vermeintliche Antinomie der Mengenlehre. Acta Mathematica*, 32:177–184, 1909.
[Schönflies, 1911] A. Schönflies. *Über die Stellung der Definition in der Axiomatik. Jahresbericht der Deutschen Mathematiker-Vereinigung*, 20:222–255, 1911.

[Schröder, 1890] E. Schröder. *Vorlesungen über die Algebra der Logik, Bd.1.* Teubner, Leipzig, 1890.
[Schütte, 1953] K. Schütte. *Zur Widerspruchsfreiheit einer typenfreien Logik. Mathematische Annalen*, 125:394–400, 1953.
[Schütte, 1954] K. Schütte. *Ein widerspruchfreies System der Analysis auf typenfreier Grundlage. Mathematische Zeitschrift*, 61:160–179, 1954.
[Schütte, 1960] K. Schütte. *Beweistheorie.* Springer, Berlin-Heidelberg, 1960.
[Schütte, 1977] K. Schütte. *Proof Theory.* Springer, Heidelberg, 1977.
[Scott, 1972] D. Scott. Continuous lattices. In W. Lawvere, editor, *Toposes, Algebraic Geometry and Logic*, volume 274 of *Springer Lecture Notes in Mathematics*, pages 97–136, 1972.
[Scott, 1975] D. Scott. Combinators and classes. In C. Böhm, editor, *Lambda-calculus and Computer Science*, volume 37 of *Springer Lecture Notes in Computer Science*, pages 1–26, 1975.
[Seldin, 1976] J. P. Seldin. Recent advances in Curry's program. *Rendiconti del Seminario Matematico dell'Università e del Politecnico di Torino*, 35:77–88, 1976.
[Seldin, 1980] J. P. Seldin. Curry's program. In Hindley and Seldin [1980], pages 3–33.
[Seligman and Moss, 1997] J. Seligman and L. S. Moss. *Situation Theory.* In Benthem and Meulen [1997], pages 239–310.
[Sheard, 1994] M. Sheard. A guide to truth predicate in the modern era. *Journal of Symbolic Logic*, 59:1032–1054, 1994.
[Shen-Yuting, 1953] Shen-Yuting. The paradox of the class of all grounded sets. *The Journal of Symbolic Logic*, 18:114, 1953.
[Shirahata, 1996] M. Shirahata. A linear conservative extension of Zermelo-Fraenkel Set Theory. *Studia Logica*, 56:361–392, 1996.
[Sieg, 1999] W. Sieg. *Hilbert's Programs: 1917-1922. The Bulletin of Symbolic Logic*, 5:2–43, 1999.
[Simmons, 1988] K. Simmons. On a medieval solution to the liar paradox. *History and Philosophy of Logic*, 8, 1988.
[Simmons, 1993] K. Simmons. *Universality and the Liar. An Essay on Truth and Diagonal Argument.* Cambridge University Press, Cambridge, 1993.
[Simpson, 1999] S.G. Simpson. Subsystems of second order arithmetic. *Perspectives in Mathematical Logic*, 1999.
[Skolem, 1957a] T.S. Skolem. *Bemerkungen zum Komprehensionsaxiom. Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 3:3–17, 1957.
[Skolem, 1957b] T.S. Skolem. *Mengenlehre gegründet auf einer Logik mit unendlich vielen Wahrheitswerten. Sitzungsberichte Berliner Mathematische Gesellschaft*, 58:41–56, 1957.
[Skolem, 1957c] T.S. Skolem. Two remarks on set theory. *Mathematica Scandinava*, 5:40–46, 1957.
[Skolem, 1960] T.S. Skolem. A set theory based on a certain three valued logic. *Mathematica Scandinava*, 8:127–136, 1960.
[Skolem, 1963] T.S. Skolem. Studies on the axiom of comprehension. *Notre Dame Journal of Formal Logic*, 4:162–170, 1963.
[Smith, 1984] J. Smith. An interpretation of Martin-Löf's type theory in a type-free theory of propositions. *Journal of Symbolic Logic*, 49:730–753, 1984.
[Smith, 1987] S. Smith. Non standard characterization of recursive saturation and resplendency. *Journal of Symbolic Logic*, 52:842–863, 1987.
[Smorinski, 1981] C. Smorinski. Fifty years of self-reference. *Notre Dame Journal of Formal Logic*, 1981.
[Smorinski, 1985] C. Smorinski. The development of self-reference: Löb's theorem. In *[Drucker, 1991]*, pages 110–133, Springer, Berlin-Heidelberg, 1985.
[Smullyan, 1957] R. Smullyan. Languages in which self-reference is possible. *Journal of Symbolic Logic*, 22:55–67, 1957.
[Sobocinski, 1949] B. Sobocinski. L'analyse de l'antinomie russellienne par Lesniewski. *Methodos*, 1:94–107/220–238/308–316, 1949.
[Specker, 1962] E. Specker. Typical ambiguity. In E. E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology and Philosophy of Science*, pages 116–123, Stanford, 1962. Stanford University Press. Proceedings of the 1960 International Congress.
[Specker, 2001a] E. Specker. *Die Antwort von Finsler.* In Gasser and Volken [1975], pages 50–56.

[Specker, 2001b] E. Specker. *La réponse de Dmitry Mirimanoff*. In Gasser and Volken [1975], pages 44–49.
[Spector, 1961] C. Spector. Inductively defined sets of natural numbers. In *Infinitistic Methods*, pages 97–102. Pergamon Press, New York, 1961.
[Surma *et al.*, 1992] S. Surma, J. Srzednicki, D. Barnett, and V. F. Rickey, editors. *St. Lesniewski. Collected Works, 2 Voll.*, volume 2. Kluwer, Dordrecht, 1992.
[Takahashi, 1989] S. Takahashi. Monotone inductive definitions in a constructive theory of functions and classes. *Annals of Pure and Applied Logic*, 42:255–279, 1989.
[Talcott and Mason, 2002] C. Talcott and J. Mason. *Feferman-Landin Logic*. In W. Sieg, R. Sommer, and C. Talcott, editors, *Reflections on the Foundations of Mathematics*, pages 293–328, Natick, Mass., 2002. Association of Symbolic Logic, A. K. Peters.
[Talcott, 1992] C. Talcott. A theory of program and data type specification. *Theoretical Computer Science*, 104:129–159, 1992.
[Tarski, 1931] A. Tarski. Sur les ensembles définissables des nombres réels vol.1. *Fundamenta Mathematicae*, XVII:210–239, 1931.
[Tarski, 1936] A. Tarski. *Der Wahrheitsbegriff in den formalisierten Sprachen*. *Studia Philosophica*, 1:261–405, 1936.
[Tennant, 1982] N. Tennant. Proof and paradox. *Dialectica*, 36:265–296, 1982.
[Tennant, 1995] N. Tennant. On paradox without self-reference. *Analysis*, 55.3:199–207, 1995.
[Thomason, 1980] R. Thomason. A note of syntactical treatments of modality. *Synthèese*, 44:391–395, 1980.
[Troelstra and van Dalen, 1988] A. Troelstra and D. van Dalen. *Constructivism in Mathematics I,II*. Elsevier-North Holland, Amsterdam, 1988.
[Turner, 1983] R. Turner. Montague semantics, nominalization and Scott's domains. *Linguistics and Philosophy*, 6:259–288, 1983.
[Turner, 1987] R. Turner. A theory of properties. *Journal of Symbolic Logic*, 52:455–471, 1987.
[Turner, 1990] R. Turner. *Truth and Modality for Knowledge Representation*. Pitman, London, 1990.
[Turner, 1997] R. Turner. Types. In Benthem and Meulen [1997], pages 535–586.
[Ursini, 1977] A. Ursini. A sequence of theories for arithmetic whose union is complete. *Rendiconti del Seminario Matematico dell'Universitá di Padova*, 57:75–92, 1977.
[Ursini, 1978] A. Ursini. On the set of 'meaningful' sentences of arithmetic. *Bollettino dell'Unione Matematica Italiana, V, serie A*, 16:237–241, 1978.
[van Dalen and Monna, 1972] D. van Dalen and A. Monna. *Sets and Integration. An Outline of the Development*. Wolters-Nordhoff, Groningen, 1972.
[van Dalen, 1999] D. van Dalen. *Mystic, Geometer and Intuitionist: the Life of L. E. J. Brouwer*. Clarendon, Oxford, 1999.
[van Fraassen, 1966] B. van Fraassen. Singular terms, truth-value gaps and free logic. *Journal of Philosophy*, 63:481–494, 1966.
[van Fraassen, 1968] B. van Fraassen. Presupposition, implication and self-reference. *Journal of Philosophy*, 65:135–152, 1968.
[van Fraassen, 1970] B. van Fraassen. Truth and paradoxical consequences. In R.L. Martin, editor, *The Paradox of the Liar*, pages 13–24. Ridgeview Pu. Co., Atascadero CA, 1970. (2nd edition 1978).
[van Fraassen, 1971] B. van Fraassen. *Formal Semantics and Logic*. The Mac Millan Press, New York, 1971.
[van Fraassen, 1972] B. van Fraassen. Inference and self-reference. In D. Davidson and G. Harman, editors, *Semantics of Natural Language*, pages 695–708, Reidel, Dordrecht, 1972.
[van Heijenoort, 1967] J. van Heijenoort. *From Frege to Gödel. A Source Book in Mathematical Logic 1879-1931*. Harvard University Press, Cambridge, Mass., 1967. (3rd printing 1977).
[Visser, 1984] A. Visser. Four-valued semantics and the liar. *Journal of Philosophical Logic*, 13:181–212, 1984.
[Visser, 1989] A. Visser. Semantics and the liar paradox. In Gabbay and Guenthner [1989], pages 617–706.
[von Neumann, 1923] J. von Neumann. Zur Einführung der transfiniten Zahlen. *Acta Litterarum ac Scientiarum Regiae Universitatis Hungariae Francisco Josephinae, Sectio Scientiarum math. 1*, 1:199–208, 1923.
[von Neumann, 1925] J. von Neumann. *Eine Axiomatisierung der Mengenlehre*. *Journal für die reine und angewandte Mathematik*, 154:219–240, 1925. Corrections in vol.155, 1926, p. 128.

[Vopenka, 1966] P. Vopenka. A new proof of the Gödel's result on non-provability of consistency. *Bulletin de l'Académie Polonaise des Sciences*, 14:11–115, 1966.
[Wagon, 1985] S. Wagon. *The Banach-Tarski Paradox*. Cambridge University Press, Cambridge 1985.
[Wang, 2001] H. Wang. *A Logical Journey. From Gödel to Philosophy*. The MIT Press, Cambridge, Mass., 2001.
[Welch, 2001a] P. Welch. On Gupta-Belnap revision theories of truth, Kripkean fixed points and the next stable. *The Bulletin of Symbolic Logic*, pages 1–15, 2001.
[Welch, 2001b] P. Welch. On revision operators. *submitted to Journal of Symbolic Logic*, 2001.
[Weydert, 1988] E. Weydert. *How to approximate the naive comprehension scheme inside classical logic*. PhD thesis, Rheinische Friedrich-Wilhelms-Universität, Bonn, 1988.
[Weyl, 1910] H. Weyl. *Über die Definitionen der mathematischen Grundbegriffe. Mathematisch-naturwissenschaftliche Blätter*, 7:93–95; 109–113, 1910.
[Weyl, 1918] H. Weyl. *Das Kontinuum*. Veit, Leipzig, 1918.
[Weyl, 1921] H. Weyl. *Über die neue Grundlagenkrise der Mathematik*. *Mathematische Zeitschrift*, 10:37–79, 1921.
[White, 1979] R.B. White. The consistency of the axiom of comprehension in the infinite value predicate logic of Lukasiewicz. *Journal of Philosophical Logic*, 8:509–534, 1979.
[White, 1987] R.B. White. A demonstrably consistent type-free extension of the logic BCK. *Mathematica Japonica*, 32:149–169, 1987.
[White, 1993] R.B. White. A consistent theory of attributes in a logic without contraction. *Studia Logica*, 52:113–142, 1993.
[Wolenski, 1995] J. Wolenski. On Tarski's background. In Hintikka [1995b], pages 331–341.
[Woodruff, 1984] P. Woodruff. Paradox, truth and logic. part I: paradox and truth. *Journal of Philosophical Logic*, 13:213–233, 1984.
[Yablo, 1982] S. Yablo. Grounding, dependence and paradox. *Journal of Philosophical Logic*, 11:111–137, 1982.
[Yablo, 1985] S. Yablo. Truth and reflection. *Journal of Philosophical Logic*, 14:297–349, 1985.
[Yablo, 1993] S. Yablo. Paradox without self-reference. *Analysis*, 53:251–252, 1993.
[Yaqūb, 1993] A.M. Yaqūb. *The Liar Speaks the Truth. A Defense of the Revision Theory of Truth*. Oxford University Press, Oxford, 1993.
[Zambelli, 1999] P. Zambelli. Alexandre Koyré alla scuola di Husserl a Gottinga. *Giornale Critico della Filosofia Italiana*, 19:303–354, 1999.
[Zermelo, 1908a] E. Zermelo. *Neuer Beweis für die Möglichkeit einer Wohlordnung. Mathematische Annalen*, 65:107–128, 1908.
[Zermelo, 1908b] E. Zermelo. *Untersuchungen über die Grundlagen der Mengenlehre I. Mathematische Annalen*, 65:261–281, 1908.
[Zermelo, 1909] E. Zermelo. Sur les ensembles finis et le principe d'induction compléte. *Acta Mathematica*, 32:186–193, 1909.
[Zwicker, 1987] W.S. Zwicker. Playing games with games: the hypergame paradox. *American Mathematical Monthly*, 94:507–514, 1987.

INDEX

$A^{\mathcal{T}}$, 209
$[x_1, ..., x_n]$ notation, 726
β-redex, reduction, 732
$\beta\eta$-reduction, 739, 749
δ, discriminator, 733
η-postponement, 739
λ notation, origin, 731
λ, short for "λ-calculus", 723
λ-calculus, 684
λI-calculus, 731, 738
λK-calculus, 731
$\mathcal{T}_n$, 209
ω, type-constant, 773
(ζ), rule, 729

4, see *Becker's axiom4*86

A (generalized abstraction in Church's system), 919
Abadi, M., 750
Abdali, S. K., 729
Abramsky, S., 785, 787, 789–791
absolute geometry, 592
absolute idealism, 4
absolutely undecidable propositions, 655–657
abstract
  classes, 65
  entity, 64, 65
  general idea, 59
  mathematical entity, 33
abstract projective geometries, 580
abstraction, 821, 835, 856, *858*, 863, *865*
  bracket, 828, 829, 835, 864
  class, 919, 924, 941, 945
  λ-, 918, 932, 937, 938, 945
  unrestricted, 947
Abstraction (or Comprehension) axiom, 9
abstraction in CL, 726
  algorithms, 728–730
  full, 784–785, 789, 791
  notations, very early, 725
abstracts, 46
$\{x \mid \varphi\}$, 947
$\{x \mid \varphi\}$, 938, 946, 949, 968, 992
$\lambda x.M$, 918, 921
absurdity, 492, 848
accidental class, 292
Ackermann function, 364
Ackermann, W., 82, 89, 90, 94, 98, 147, 172, 202, 206, 227, 233, 237, 247, 250, 251, 354, 357–359, 362, 366, 367, 378, 391, 392, 408, 583, 672, 711, 716, 728, 903, 906, 907, 926, 931–933
Aczel, P., 83, 117, 741, 781, 905, 906, 922, 927, 930, 939–941, 952, 970, 977, 984, 996, 997
Addison, J., 565, 725
adjoint functors, 756
admissible, 101
  rules, 100, 102
  rules of inference, 100
affine logic, 790
Ajdukiewicz, K., 309, 517, 752
Akama, S., 995
Alexandroff, Paul, 826
algebra of logic, 513, 525
algebraic aspects, 162
algebraic interpretation, 103
algebraic logic, 524, 526, 736
algebraic semantics, 103
algebraic tradition, 171

algebraization of logic, 522, 581
algebraization of set theory, 603
Algol language, 742, 760
algorithms for typability, 747, 774–776
all-embracing One, 72
(α), *858*
α, 211, 857
Amadio, R., 783
analysis
  intuitionistic, 485, 496, 500
analysis, consistency of, 667
analytic geometry, 528
analytic philosophy, 1, 21
analytic truth, 30
Anaphora, 389, 395, 407, 418, 430–432, 436
Anderson, C. A., 730
Andréka, H., 589, 736
Andrews, P. B., 207, 217–218, 230, 247, 248, 737, 766
Anellis
  Archives, 250
  Irving H., 195, 210, 216, 226, 231, 237, 244, 247
Annahmen, 35, 46, 58
  thesis, 58
Anscombe, E., 258
anti-foundation, 905, 977, 978, 993, 997
anti-prenex, see *direction, anti-prenex* and *form, anti-prenex*254
anti-psychologism, 5
AFA, 977–979, 993
antimonies, 314
antinomy, see *paradox*875
Antonelli, G. A., 975, 977
apartness, 81
APL language, 744
applicability, 856
application, 821, 826, 827, 856, *858*, 863, 864
application term, 917, 943, 950
applied λ and CL, 723
approximants, 749
approximate normal form, 779
Aquinas, T., 32
Arai, T., 713
Aristotelian
  inherence, 69–71
  metaphysics, 61
  realism, 72
  thesis, 72
Aristotle, 60, 61, 72, 307, 313, 529, 541, 555
arithmetic, 112
  consistency of, 667
  finitist, 355
  primitive, 368, 374
  primitive recursion, 364
  of real numbers, 513
  recursive, 363, 377
arithmetical hierarchy, 651–652
arithmetization, 336, 468
arithmetizations of analysis, 331
Artemov, S., 487
artificial intelligence, 742
Artin
  's Theorem, 238–239, 247
    Artin's proof, 238, 248
  Emil, 198, 206, 238, 239, 247, 252
Asher, N., 994
Asperti, A., 750, 989
Asser, G., 386, 404, 410, 411
assertional system, 846
assumption, 46
AT, 986
atom, 845
atomism, 263, 264, 289
Atten, M. van, 83, 96, 98, 120
Außersein, 61
Austin, J., 978
Autexier, Serge, 195, 224
Automath, 760–761, 783
  book, line, 760
automorphism
  shifting, 992

Avigad, J., 325
Ax, J., 565
axiom, 845
  of choice, 210, 214, 224–229, 396, 497, 498, 598, 600, 602, 603
    weak forms, see *Principle of Dependent Choice, Kœnig's Lemma*254
  of infinity, 206
  of reducibility, 206
  of substitutability, 230
axiom of constructibility, 497
axiom of countable choice, 89
axiom of elementary sets, 600
axiom of extensionality, 600
Axiom of inaccessible cardinals, 601
axiom of inaccessible set, 602
axiom of infinity, 161
axiom of reducibility, 11
axiom of regularity, 153
axiom of replacement, 153, 154, 602
axiom of separation, 153, 600
axiom system, 515
axiomatic method, 328, 519, 521, 522
axiomatic set theory, 152, 176
axiomatics
  contextual, 360
  existential, 326, 355, 361, 371, 377
  formal, 360
  sharpened, 377
axiomatisability, 514–516
axioms, 40, 50, 57
  Dedekind–Peano, 230
  of equality, 230, 232
  of set theory, 44

B-H-K interpretation, 89
Börger, E., 175
Backus, J., 744
Badesa, C., 83, 172, 174, 176, 177
Baer, R., 178
Baire, R., 888
ball, 595
Banach, S., 596, 597
Banach-Tarski paradox, 596, 597
bar induction, 116
bar theorem, 78, 113, 114, 492, 496
Bar-Hillel, Y., 755
Barendregt, H. P., 723, 725, 730, 741, 744–746, 749, 773, 779–780
  ω-rule, 780
  quasi-leftmost reductions, 749
Barr, M., 790
Barwise, J., 967, 976–979, 997
Barzin, M., 82, 483
Bays, T., 178
Bealer, G., 997
Becker's axiom (4), 486
Becker, O., 483, 486
Beeson, M., 83, 117, 950, 955
Beethoven, 70, 71
Behboud, A., 348
Behmann, H., 346, 500, 727, 826, 918, 926, 927, 931, 952
Behnke, H., 348
being, 35, 63
being thus-and-so, 33, 35
being-independent being-indifferent *So-seine*, 64
beingless impossible Meinongian object, 70
beingless incomplete objects, 61
beingless Meinongian objects, 64
beingless objects, 64–66, 68–71
beingless One, 72
beinglessness, 68
Bell, J. L., 179, 992
Belnap, N., 967, 973–975, 994
Ben-Yelles, C., algorithm, 776
Bénabou, J., 762
Benacceraf, P., 414
Benaissa, Z., 750
Benzmüller, Christoph, 195
Berarducci, A., 746, 769
Bergstra, J., 746, 749
Berkeley, G., 32, 37, 59
Bernardi, C., 946, 988

Bernays, P., 84, 86, 87, 93, 104, 135, 152, 172,198, 202, 206, 212, 214, 217–218, 220, 222, 227, 231–233, 237, 247, 250, 321, 323, 346, 354, 355, 385, 387, 388, 396, 402, 407, 418, 434, 436, 485, 489–492, 495, 519, 522, 573, 619, 620, 660, 667, 670, 672, 675, 682, 700, 703, 706, 707, 709, 712, 727, 728, 736, 743, 826, 895–897, 921, 932, 945, 989
Bernstein, F., 353, 878, 887, 888, 894
Berry, G., 536, 785, 786, 791, 895
Berry's Paradox, 442, 443, 536
(β), *858*
β, 211, 857
Beth model, 106, 116, 487, 489
Beth semantics, 488
Beth, E. W., 83, 106, 108, 117, 565, 594, 700
Beth–Kripke semantics, 103
Beweistheorie, 322
Bezem, M., 83
Birkeland, K., 128
Birkhoff, G., 103, 172, 573, 825
Bishop, E., 84, 109, 117
bisimulation, 906, 979
Blass, A., 790, 791
Blok, W. J., 522, 524, 526, 608
Bloo, R., 750
Blue Mosque, Istanbul, 30
Bochvar, D. A., 924–926, 959, 962
Boffa, M., 992, 994
Böhm, C., 741, 743–747, 769, 772, 777, 780
  Böhm's theorem, 745, 747
  compiler, 743
  numerals, 746
  range theorem, 746
Böhm tree, Böhming out, 745, 780
Bolyai-Lobachevsky geometry, 592
Bolzano, B., 141, 337, 449, 530
Boole, G., 139, 314
Boole–Peirce–Schröder tradition, 171
Boolean algebra, 523–526, 545, 563, 581–584, 586, 587, 589
Boolean clones, 633–641
Boolean equations, 163
Boolean functions, 633–641
Boolean operation, 539, 540
Boolean set algebra, 524
Boolean values, 769
Boolos, G., 968, 989–991
Boos, W., 177
Borel, E., 82, 353, 888, 892, 894, 895, 934, 945
Born, M., 338, 345
Bostock, D., 413, 414
bound variables in LISP, 742
Bourbaki, N., 386, 395, 396
Bradley's regress, 260
Bradley, F. H., 3
Brady, G., 172
Brady, R. T., 956, 958
Brentanian, 64
  Aristotelianism, 72
  empiricism, 59
  intentionality thesis, 33
  thesis, 30
Brentano's intentionality thesis, 30, 34
Brentano, F., 29, 32, 60, 61, 314
Briaud, D., 750
Bridgeman, P. W., 825
Bridges, D., 84
Broda, S., 776
Brouwer and intuitionism, 168, 180
Brouwer's bar theorem, 113
Brouwer's continuity principle, 112
Brouwer's fan theorem, 112
Brouwer's fixed point theorem, 78
Brouwer's influence on Gödel, 96
Brouwer's principle, 114
Brouwer's principle for functions, 114
Brouwer's theory of species, 83
Brouwer, L. E. J., 77–81, 84, 87, 89, 92, 100, 105, 108–113, 115,

117, 200, 204, 235, 248, 324, 348, 350, 353, 354, 490, 492, 493, 495–502, 607, 683, 888, 892, 894
Brouwer–Heyting–Kolmogorov (B-H-K) explication, 88
Brouwerian algebra, 581
Brouwerian assertion of p, 82
Brouwerian counterexample, 92
Bruce, K., 783
de Bruijn, N. G., 757, 762
  Automath, 760–761, 783
  variable-less λ notation, 750
Brun, V., 128, 179
Buchholz, W., 711
Bunder, M. W., 741, 776, 840
Burali-Forti, C., 8, 725
Burge, T., 979
Burge, W. H., 743
Burgess, J., 966, 967, 972, 974, 975, 991
Burris, S., 173
Bussotti, Paolo, 195, 244

C-system, 840
Cain, J., 965
calculus ratiocinator, 23
calculus of (binary) relations, 586
calculus of classes, 545, 547, 549, 562, 581, 582, 586
calculus of constructions, 772
calculus of deductive systems, 523, 525, 581, 586
calculus of relations, 583, 584, 586, 606
calculus of sentences, 523, 524, 563, 581, 586
calculus of systems, 525, 526, 563
call by value, 743
Campbell, R., 64
canonical elements, 765
canonical real number generator, 111
canonical terms, 834
Cantini, A., 944, 955, 965–968, 982, 985, 993, 995
Cantor's absolute, 601, 603
Cantor's Paradox, 441, 442
Cantor's second number class, 79, 109
Cantor's set theory, 79
Cantor's theorem, 881, 883, 885, 896, 901
Cantor, G., 8, 15, 80, 109, 131, 153, 316, 325, 337, 338, 340, 353, 441, 442, 499, 878–881, 883, 885, 886, 888, 893, 897, 905, 988
Cardelli, L., 750, 774
cardinal addition, 603, 604
cardinal algebra, 603–605
cardinal arithmetic, 13, 14, 598, 603, 604
cardinal degree of completeness, 516, 518
cardinal number, 600, see *number, cardinal*875
cardinal numbers, 598
Cardone, F., 774
Carlisle, T., 997
Carnap, R., 498, 561, 714, 910–912, 918, 921, 924, 980
Carroll, L., 314
Cartesian closed category, 755–757, 777, 785
Cartesian product, 41
Cartesian space, 592
Cassirer, E., 348
categorial grammar, 752, 755
categorical, 536
categorical abstract machine, 786
categorical combinators (Curien), 786
categorical set, 534, 536
categorical set of sentences, 536
categoricity, 156
category theory, 952, 992
category theory and types, 755–759
CFA, 939
champ fini, 208–209, 214–215, 220–

221, 237
definition, 209
Chang, 564, 565, 568, 957
Chang, C. C., 160, 956
Chapuis, A., 967, 987
characteristica universalis, 23
characteristica universalis, 498
Chauvin, A., 741, 961
Chevalley
Catherine, 197
Claude, 197, 204, 207, 245, 246, 248, 369
Chierchia, G., 997
Chin, L., 584, 589
Chisholm, R. M., 36, 62
choice, 45
choice law, 110
choice sequences, 109, 116, 488, 495–497
Chu, P. H., 790
Church's rule, 105
Church's thesis, 104, 105, 114, 628, 631, 643, 644, 732, 791, 823
Church, A., 82, 95, 98, 171, 175, 217, 252, 309, 374, 379, 483, 497, 575, 576, 618, 619, 624, 628, 631, 632, 641–644, 649, 655, 657, 660, 684, 714, 729–733, 737–738, 746, 749, 819, 856, 860, 909, 917–924, 929, 931, 953, 970
Church numerals, 731
Church-Rosser, see *confluence*732
Church-Rosser Theorem, 823, *860*, 865
Chwistek, L., 307, 737, 751, 903, 907–909, 911, 912, 929, 934
circular class, 928
circular concept, 905, 971, 975
circular definition, 890, 903, 905, 975
circular predicate, 975
circularity, 891–893, 900, 903, 907, 910, 942, 977, 987, 997
CL, for "combinatory logic", 723
class
full satisfaction, 985, 986
non-extensional partial, 947
partial, 947
proper, 967
total, 947, 953
class of all ordinals, see ON875
classical decision problem, 175
classical deduction theorem, 50
classical deductive systems, 516
classical first-order arithmetic, 77
classical logic, 77, 684, 848
classical second-order logic, 77
closure
reflective, 983–985
schematic, 984
weak, 983
semantical, 934, 959, 961, 968, 980
transitive, 924
Cocchiarella, N., 385, 993
cofinality, 601
cofinality of leftmost reductions, 749
quasi-leftmost, 749
Cogan, Edward J., 839
Cohen's independence proofs, 158
Cohen, P. J., 158, 497, 498, 500
coherence space, 786
coincidence, 81
coinduction, 979
combinator, 826
B, 826, *863*
basic, 863
C, 826, *863*
commutator, *863*
composition operator, *863*
duplicator, *863*
fixed point, *859*
I, 822, 826, *863*
J, 822, 863
K, 826, *863*
S, 826, *863*
W, 826, *863*
Y, *859*
combinatorial completeness, 921, 922

combinatorially complete, 726, 728
combinatorics, 132, 180
combinators, 917, 924, 934, 936, 937, 954
combinators, basic, 726, 729
combinatory algebra, 781, 938, 939, 942, 946, 954, 968
  effective, 944
  extensional, 946
  partial, 951
combinatory completeness, 826, 828
combinatory logic, 820, 822, 863, *864*
  illative, 836, 840
  pure, 836
combinatory reduction systems, 738
compact cardinal, 602
compactness
  proof, 35
  theorem, 54
compactness problem, 602
compactness theorem, 602
compiler, Böhm's, 743
complement of a property, 40
complemented ordered lattices, 580
complete, 516, 534, 536, 540
complete developments, 738
complete existent object, 62
complete existent or subsistent object, 61
complete object, 61
complete systems, 526
completeness, 35, 49, 103, 106–108, 202, 206, 227–230, 330, 514–516, 520, 529, 536, 547, 554, 567, 572
  definition, 202
  semantical, 482, 487
completeness of concepts, 534
completeness of first-order logic, 174
completeness problem, 592
completeness theorem, 366
  intuitionistic, 482, 488, 489
  classical, 449
comprehension, 46
  elementary (ECA), 954, 955
  positive
    non-uniform, 993
    uniform, 994
  stratified (SCA), 954
  type-free, 876, 930, 938, 955, 956, 958, 966, 968
  uniform, 954
  uniform ($\mathcal{F}$-CA), 994
  untyped, 924, 934, 946, 995
comprehension principle, 58
$Comp(\mathcal{A})$, 944
computable functionals, 482, 489–491, 493–495, 754
computable functions, 490, 491, 493, 494, 732
computable reals, 887
concept
  of being, 67
  of set, 45
concrete domains, 785
condensed detachment, 775, 776
conditional, 38, 42
conditional expressions, 742
confluence, 732, 738, 748
  of β-reduction, 732, 748, 750
  of weak reduction, 732, 748
confluence theorem, 823
conjunction, 42
connectives, 38
consequence, 514, 523, 525, 536, 547
consistency, 35, 49, 58, 114, 156, 200–202, 333, 338, 514–516, 520, 529, 547, 554, 567, 914, 955, 956, 958, 969, 980, 985, 992–994
  of arithmetic, 202, 230–233, 235–238
  of first-order logic, 199, 226
  proof of, 202, 230–233, 235–238
    Gentzen's, see *Gentzen's consistency proof*254
    Herbrand-style, see *Herbrand-style consistency proof*254

semantic, 333
syntactic, 345
consistency of arithmetic, 680, 703
consistency problem, 321, 325, 332, 337, 341, 343
consistency program
finitist, 324, 347
consistency proof, 322, 367, 376
of PA, 484, 485, 490, 491, 493
direct, 342, 344, 356, 362
consistency theorem, 363, 366
Constable, R. L., 765, 788
constant, *858*, 864
constants, 38
constitutive property, 33, 37, 62–64, 66, 70, 71
constructions, calculus of, 772
constructive disjunction, 98
constructors, 765
content, 29
contexts, 747, 862
contextual equivalence, 747
continuations, 779
continuity, 330, 332
continuity (completeness), 328
continuity principle, 78, 113, 497, 498
continuous lattices, 777, 787
continuum, 108, 115, 498
continuum hypothesis, 158, 655–656, 659
contraction, 857, *859*
α, *859*
β, *859*
η, *859*
contraction, contractum, 732
contractum, 859, *859*
contradiction, see *paradox*875
convention T, 547, 549
convention T, 551, 552
conversion, 819, 826, *859*
β, *859*
$=_\beta$, *859*
$\beta\eta$, *859*
$=_{\beta\eta}$, *859*
$=_*$, *859*
weak, 863, *865*
$=_w$, 863, *865*
for λ-terms, 866
conversion-invariant, 841
Conway, J. H., 790
Copi, I., 388, 438, 439, 443
Coppo, M., 773–774, 787
Coquand, T., 764, 787, 791, 989
calculus of constructions, 740, 772
Corcoran, J., 522, 530, 545, 546, 608
corecursion, 979
correction (of Herbrand's False Lemma)
Gödel's and Dreben's, 217–218, 221
definition, 218
Heijenoort's, 218, 220, 221, 224
definition, 218
correlation law, 110, 117
correspondence theory of truth, 555, 558
Corry, L., 325
Count terms, 386, 413, 414
countable models, 134, 156
counteraxiom, 848
counterexample
strong, 500
weak, 500
Counterparts, 429, 430, 435
Courant, Richard, 198
Cousineau, G,, 786
Couturat, L., 499, 881, 889
CPC (classical propositional logic), 482, 483
cpo (complete partial order), 783, 787
CQC (classical predicate logic), 482, 484
Crabbé, M., 992
Craig's Interpolation Theorem, 238
Craig, W., 238, 248, 565, 736
Cresswell, M. J., 386, 417, 419, 428
Critical formulas, 390, 391
Crossley, J. N., 725
CUCH language, 744

Curien, P. L., 750, 785, 786, 791
Curry semantics, 845
Curry's paradox, 160, 733, 832
Curry, H. B., 683, 699, 723, 725, 726, 728–730, 733–735, 739, 740, 742, 749, 752, 819, 825, 856, 917, 918, 920–924, 929–931, 939, 941, 942, 956
  normalization proof, 766
  typability algorithm, 775
  undecidability theorem, 745
Curry-Howard isomorphism, 735, 838
currying, 726, 756, 856
cut, 669
cut elimination, 687, 688, 931, 933, 949, 986, 995, 996
Cut Elimination Theorem, 834
cycle, 926, 928, 973, 975
cylindric algebra, 586–589, 736
cylindric set algebra, 586–588

D_∞, *see* models of untyped λ
D'Agostino, G., 988
van Dalen, D., 83, 96, 117, 236, 248, 724, 746, 760
Damas, L., 775, 776
Damnjanovic, Z., 965
Dantzig, D. van, 97
Daumal, R., 37, 41
David, M., 556
Davidson, D., 559
Davies, N., 997
Davis–Putnam procedure, 242
Davis, M., 175, 241–242, 248, 628, 653, 659–662, 732
Davison, 559
Dawson, C., 708
Dawson, John W., Jr., 199, 235, 244, 248, 249, 708
De Morgan, A., 314
De Rouilhan, P., 891
de Swart, H. C. M., 489
decidability, 566, 567, 572, 575, 581, 589
decidability of elementary Euclidean geometry, 590
decidability of elementary geometry, 573
decidability of the arithmetic of the natural numbers with addition, 575
decidability of the first-order theory of the reals, 567, 573
decidability of the theory of reals, 566
decidability problem, 592
decidable, 540
decidable classes, 146
decidable theories, 151
decision domain, 518
decision method, 573
decision method for real-closed fields, 567
decision method of eliminating quantifiers for axiomatic theories, 134
decision problem, 134, 566, 586, 593
decision procedure, 566, 568, 570, 571
decision procedure for the elementary theory of multiplication of natural numbers, 152
decision procedure for the first-order theory of reals, 571
Dedekind, R., 79, 80, 200, 321, 325, 328, 330, 332, 333, 335–343, 353, 360, 379, 499, 600 see also *axioms, Dedekind–Peano*254
deduction theorem, 57, 58, 517, 821, 829
deductions as terms, 735, 759, 765
deductive method, 529
deductive system, 515, 525, 547
deductive systems (Lambek), 758
deductive theory, 529
deductively valid, 40
deep, see *inference, deep*254
definability, 530–532, 534, 536–538, 564, 565, 572, 607
definability of the reals, 571

definability, absolute notion of, 657–659
definable concept, 531
definable set, 538, 539
definable set of real numbers, 539, 540
definable set of reals, 540
Definite Descriptions, 6, 385, 387, 416, 419, 430, 432, 436
definite eigenschaften, 177
definite property, 134, 152
definite set, 845
definition, 581
definition of truth, 550
definitions, 40, 307
deflationary theory of truth, 555, 558
degrees of unsolvability, 645–648, 650–654
Dekkers, W., 741
δ, 211
δ-quantifier, see *quantifier, δ*-254
δ-variable, see *variable, δ*-254
denial
  external, 978
denumerably unfinished, 499
denumerably unfinished set, 79, 109
dependent types, 740, 761, 764, 771
derivability, 532
derivability condition, 370
derivable, 100, 101
derivable sentence, 531
Desargues' axiom, 593
Descartes, R., 81
descriptors, 46
designated reference lcass, 41
designation, 41
detachable subspecies, 115
Detlefsen, M., 392
developments, 738
  equivalence theorem, 749
  finiteness theorem, 739, 749
  minimal complete, 748
Dezani, M., 744, 773–774, 787
diagonal argument, 551
diagonalization, 499
diagonalization theorem, 911, 980
Dialectica, see *Gödel, K.*744
Dialectica Interpretation, 449, 485, 491, 493–496, 500
Dieudonné, Jean, 197, 207, 248
dilator, 786
Diller, J., 83, 767
D_∞, 945
Diophantine equations, 131
Diophantine sets, 175
Diophantische Gleichungen, 175
direct proof, 334
direction
  anti-prenex, 214
    definition, 213
  prenex, 214
    definition, 213
disjunction, 105, 118
disjunction property, 98, 100, 101
disjunctive normal form, 622, 633–634
Diskin, Z. B., 781
dispute with Zermelo, 178
disquotational theory of truth, 556, 557
distributive lattices, 581
Dixon, A. C., 887
D_κ, 945, 946
doctrine of external relations, 5
doctrine of internal relations, 4, 5
domain, 41, 46, 856
  concrete, 785
  effective (Ershov), 788
  qualitative, 786
  reflexive, 778
  Scott, 777, 786–787
  universal, 783
domain comprehension principle, 69
domain theory, 783–789
dominance, 788
Doner, J., 575, 608
Donnellan, K., 416–418, 434, 444
double negation shift, 114

Dragalin, A., 83, 701, 767
Drakengren, T., 783
Dreben, B., 174, 175, 203, 210, 217, 218, 237, 245, 246, 248, 253
Dreben–van Heijenoort introductory note, 175
Dries, L. van den, 176, 575, 608
Droste, M., 783
Dummett, M. A. E., 83, 173, 488, 684
Dunn, M., 967
Dybyg, K., 742
Dyckhoff, R., 701
dynamic binding, 742
Dyson, V., 117
Dzik, W., 100

ε-substitution method, 358
ε-theorem
  first, 365
  second, 366
Ebbinghaus, H.-D., 176, 342
Eckart, C., 71
Edelberg, W., 429, 430
Edinburgh Logical Framework, 753
effective topos, 771, 788
EFL*, 946
Egli, H., 788
Ehrenfeucht, A., 565
Eklund, M., 173
elementary algebra, 567
elementary equivalence, 565
elementary formal system, 846
elementary geometry, 593
elementary linear set, 540
elementary measure theory, 13
elementary predicate, 845
elementary propositions, 260
Elimination Theorem, 848
embedding-projection pair, 778
empirical phenomena, 72
empiricist methodology, 61
Enderton, H. B., 730
Engeler, E., 782
ENIAC, 835
entity, 41
Entscheidungsproblem, 624, 635
Entscheidungsproblem, 132, 323, 727, 732
epistemic limitatin, 72
epistemic logicism, 18
ϵ_0, 491
Epsilon Theorem, 176
epsilon-substitution method, 176
Epstein, R., 902
equality, 89, 91
equation-invariant, 841
equational theories of relation algebra, 581
equational theories of relation algebras, 580
equational theory, 580
$\equiv$, *859*
equivalence, 524
equivalence, contextual, 747
Ermine, F., 746
Errera, A., 82, 483
Ershov, Y. L., 565, 788
Essay on the Foundations of Geometry, 4
essential class, 292
essentially richer, 534, 536
essentially undecidable, 473, 576, 579, 580
essentially undecidable theory, 576, 577
Esser, O., 994
(η), *859*
η, 857
Etchemendy, J., 559, 562, 563, 608, 967, 977–979
Euclid, 3, 530
Euclidean affine geometry, 594
Euclidean geometry, 534, 573, 590, 592–595, 901
Euclidean space, 597
Evans, G., 432, 434
Ewald, P. P., 325
exclusion, 41

existence, 35, 98
existence property, 98, 105, 118
existence–subsistence distinction, 32
existenital axiomatics, 324
existent, 41, 72
  entity, 65
  object, 32–34, 41, 70, 71
existential claims, 22
existential import, 41
experimental psychology, 29
explicit substitution, 750
explicitly definable notion, 900, 901
exponentiation and λ-abstraction, 756
(Ext), 859
extended Church's thesis, 106
extensional equality, 729, 901, 903
extensional semantic domain, 69
extensionalist
  logic, 31, 35
  semantic ontologists, 66
extensionality, 312, *859*, 912, 915, 936, 941, 942, 946, 949, 958, 992–995
extensions, 46
extralogical, 61
  applications, 50
extranuclear, 37, 58
  (nonconstitutive) property, 34
  incompleteness, 37
  property, 36, 37, 58
extraontic domain commitment, 69
extraontological
  economy, 69
  jungle, 68, 69
extraontology, 65, 72

F, see *types, Curry's*734
$\mathbb{F}$ (Girard's system), 768
F-deductions, 837
$\mathcal{F}_1$, 834
$\mathcal{F}_2$, 834
$\mathcal{F}_3$, 834
fan, 111
  induction, 116
  theorem, 119
Feferman, S., 181, 237, 239, 249, 325, 378, 565, 713, 741, 900, 926, 927, 930, 939, 940, 947, 949, 951–955, 958, 961, 963, 966, 968–971, 983, 984, 986, 992, 996, 997
Felscher, W., 789
Fenstad, J. E., 160, 956
Fermüller, Christian G., 212, 215, 218, 244, 247
Ferreiròs, J., 325
Ferreira, F., 378, 992
Feys, R., 683, 723, 725, 739
Field, H., 558, 987
filter models, 773, 787
FIM, 113
Findlay, G. D., 755
Findlay, J. N., 37, 62, 64
Fine, K., 406
finitary mathematics, 490, 500
finitary reasoning, 493
finitary spread, 111, 116
finitary standpoint, 485, 494
finite axiomatization, 152
finitely axiomatisable, 579
finitely axiomatizable, 593, 594
finitely definable countable ordinals, see FOD875
finiteness of developments, 739, 749
finitest mathematics, 357
finitiesm, 367
finitiest mathematics, 373
Finitism, 390–395, 409, 415, 441
finitism, 199–202, 204–206, 224–228, 235, 323, 353, 371, 376, 379, 495
finitist mathematics, 351, 354, 376
finitist methods, 373
finitist number theory, 363
finitist proof theory, 354
finitist standpoint, 361, 375, 376
Finsler, P., 903, 905, 906, 910
first act of intuitionism, 77

first and second incompleteness theorems, 98
first-order systems, 173
Fitch, F. B., 735, 740, 741, 744, 760, 923, 924, 930, 933–939, 942, 943, 945, 947, 951, 961, 963, 969, 995, 998
Fitting, M., 83, 400, 417, 420, 422, 426, 427, 964, 968, 997
fixed point
  combinatory, 939
  intrinsic, 965
  maximal, 963, 965
  minimal, 964, 965
  optimal, 965
  set theoretic, 939
  Tarski's, 923, 964
fixed point
  Brouwer's, 957
fixed point theorem, 922, 923, 937, 941, 964, 987, 988, 995
Fixed-Model theorem, 969, 970
fixed-point combinators, 732–733, 778
fixed-point induction, 742
fixed-point, least, 747, 754
  in D_∞, 778
Flagg, R., 934, 942–946, 966, 996
FOD, 886
forcing, 107, 116, 497
forcing relation, 106
form, 261
  anti-prenex, 216
    definition, 213
  functional, 210
  prenex, 209, 216–218, 220, 221, 223, 224, 229
    definition, 213
  Skolem normal, 229
  (outer) Skolemized, 210, 212, 214–218, 221, 227, 230
    definition, 212
    example, 210
    vs. Skolem normal form, 229
  inner Skolemized, 212, 218
    definition, 212
formal definability, 530, 531
formal object, 845
formal semantics, 41
formal system, 845
formalism, 79, 853
formation, 36
  principles, 39
formula
  Δ^1_2, 976
  Δ^1_1, 992
  elementary, 954
  generalized positive, 993
  monotonic, 949, 950
  numeric, 356
  positive, 949
  stratified, 954
formulae as types, see *propositions as types*734
formulas
  verifiable, 362
formulas-as-types, 838
formulas-as-types interpretation, 918
Forrest, P., 71
Forster, T. E., 734, 992
Forti, M., 949, 977, 993
foundational matters, 179
foundations of mathematics, 1, 7
Fourman, M., 83, 787
Fraïsse, R., 565
Fraassen, B. C. van, 71
Fraenkel, A., 131, 152
Frayne, T. E., 565
free assumption, 34, 58
Free-variable statements, 391
Frege structure, 939–942, 944–946, 966
Frege's Rule V, 8
Frege, G., 1, 2, 6, 8, 9, 12, 18, 24, 31, 41, 46, 79, 139, 171, 199–201, 205, 206, 209, 216, 249, 255–257, 260, 261, 269, 297, 307–309, 311, 313, 317, 319, 339, 342, 385, 389, 392, 393, 399, 401, 413, 414, 416, 419,

426, 431, 672, 725, 726, 731, 880–883, 885, 897, 919, 940, 941, 991, 992, 996, 997
Frege–Peano–Russell–Hilbert tradition, 171
Freudenthal, H., 499
Freyd, P., 784
Friedberg, R., 648, 654, 782
Friedman, H., 83, 99, 378, 753, 940, 946, 980–982, 984, 990
full abstraction, 784–785, 789, 791
full completeness in linear logic, 790
function, 41, 819
  calculable, 374
  index, 210
  indicial, 210
  provably total, 920
  recursive, 232–235, 237–238, 860
  Skolem, 210–212
  symbol, 46
function-realizability, 114
functional calculus, 583, 586
functional completeness, 758
functional programming, 742
functionality, see *types, Curry's*734, 834, 836, 841
  basic, 837
functionals of finite type, 841
functors, 38, 46
fundamenntal sequence, 110
FV, 858

Gödel, K., 1, 77, 83, 95, 97, 98, 102, 108, 116, 135, 173, 175, 318, 323, 353, 360, 365, 367–369, 371, 375, 378, 379, 552, 553, 559, 561, 565, 618, 619, 624, 632–633, 641–643, 655–660, 667, 671, 684, 703, 706, 708, 714, 824, 841, 90, 907, 910, 914, 916, 920, 928, 929, 942, 988, 989
Gödel numbering, 824
Gödel's Completeness Theorem, 449
Gödel's Completeness Theorem, forerunners, 458
Gödel's Speed-up Theorems, 479
Gödel-Gentzen translation, 667, 671
Gödel's completeness theorem for predicate logic, 133
Gödel's dissertation, 135
Gödel's first incompleteness theorem, 552, 630, 631, 641, 644–646
Gödel, K., 8
Gómez-Torrente, M., 562, 563
Gabriel, R. P., 742
Gaifman, H., 565, 902, 979, 987
Galileo's Paradox, 442
Gallier, J., 768
game semantics, 785, 789–791
γ, 211
γ-quantifier, see *quantifier,* γ-254
γ-variable, see *variable,* γ-254
Gandy, R., 725, 737, 791
Garciadiego, A., 878, 880
Geach, P. T., 431, 432, 930, 956
Gegenstandstheorie , 61
general deductive systems, 514
general models, 737, 753
general theory of binary relations, 526
general theory of classes, 549–552
general theory of order, 526
Generalization Theorem, 821
generalized continuity, 114
generalized continuum hypothesis, 600
generalized functionality, 740
generalized quantifier, 994
generalized recursion, 943, 975, 976
Generalized Rule
  of δ-Quantification, 213, 220–223
    definition, 213
  of γ-Quantification, 212, 218, 220–223
    definition, 212
  of γ-Simplification, 213, 220–222
    definition, 213
  of Simplification, 213, 214, 218, 220, 224

definition, 213
generation lemma, 773
generative grammar, 547
Gentzen
's calculi, 203, 224, 237
's consistency proof, 233, 237
's Hauptsatz, 203, 221, 224, 237
's verschärfter Hauptsatz, 224
Gentzen, G., 77, 83, 98–100, 195, 202, 203, 211, 233, 236, 237, 249, 254, 323, 353, 356, 357, 375, 376, 378, 392, 398, 402, 475, 485, 491, 667, 740, 764, 765, 791. 834
Gentzen-style, 99
geometry, 528, 529, 536, 596
geometry of interaction, 790
geometry of straight lines, 513
George, A., 178
Geuvers, J. H., 725, 741, 750
Ghilardi, S., 83, 101
Giannini, P., 788
Gierz, G., 787
Gilmore, P. C., 241, 743, 744, 922, 947–949, 951, 958, 961, 963, 969, 994
Ginsburg, S., 659
Giovannucci, R., 744
Girard, J.-Y., 683, 702, 750, 768–770, 786, 790, 791, 953, 958, 963, 989, 995, 996
candidats de réductibilité, 768, 772
linear logic, 786, 789
paradox, 764
system $\mathbb{F}$, 768, 786
no typability test, 775
Givant, S., 606, 736
GLC (Takeuti's system), 768
Glivenko's theorems, 87, 97
Glivenko, V., 77, 82, 83, 86–88, 96, 97, 484, 670
Göbel, R., 783
Goddard, L., 417, 420–422, 424, 428
Gödel, K., 197–199, 203, 204, 206, 210, 217, 227–229, 233–235, 238, 245, 246, 248, 249, 252, 731
Dialectica paper, 744, 766, 790
system T, 766
's Completeness Theorem, 227–229
's First Incompleteness Theorem, 199, 202
's Second Incompleteness Theorem, 199, 202, 233, 235, 238
's and Dreben's correction, see *correction*254
number, 468
Goethe, 70, 71
Göttingen, 198, 206
Goldbach conjecture, 82
Goldberg, H., 49
golden mountain, 30, 32, 33, 41–43, 65
Goldfarb's negative solution of the Gödel case with identity,
Goldfarb, Warren, 173, 175, 203, 207, 210, 217, 229, 237, 245, 246, 248, 249
Goldstein, R., 180
Gonthier, G., 750
Goodman, N. D., 757, 762
Goodstein, R. L., 438
Gordeev, L., 955, 994
Gordon, M., 754
Gottingen, 529
GPC, 993, 994
Grädel, E., 175
Gramlich, Bernhard, 244
grammar, categorial, 752
graph, 977, 979
graph models of λ, 781–782
Grassmann, H., 166
Grattan-Guinness, I., 643, 656, 660, 661, 725, 880
Gray's Elegy argument, 257
Gray, J., 325

Grayson, R., 83, 787
Grezegorczyk, A., 318
Grišin, V. N., 922, 958, 992, 995
Griffin, N., 31, 62, 880
Griffin, T., 766
Gross, W., 743, 777
Grossmann, R., 62
groups, 580
Grue, K., 946
Grundesetze der Arithemetik (*The Basic Laws of Arithemetic*), 9
Guard, J. R., 766
Guerrini, S., 744
Guillaume, M., 120
Gunter, C., 774, 783
Gupta, A., 967, 969, 971, 973, 975, 987, 994
Gurevich, Y., 175
Gödel, K., 400, 401, 408, 410, 423

Höfler, A., 29
Hájek, P., 179, 958, 996
HA (Heyting Arithmetic), 482–485, 489–491, 493, 494
Hacker, P. M. S., 64
Hadamard, Jacques S., 197–198, 208, 248, 250, 888
Hahn, H., 110, 483
Halbach, V., 967, 982, 986, 994
Halldén, S., 926
Haller, R., 36
Hallett, M., 324, 325, 337
Hallmann, Andreas, 244
Halmos, P., 736
Hamm, F., 997
Hanatani, Y., 767
Hancock, P., 765, 782
Hanf, W., 565, 602
Hankin, C., 723
Hannequin, J., 881
Hardin, T., 750
Harrington, L., 966
Harrop, R., 99
Harvard University, 825
Hasse, Helmut, 198, 203, 207, 238, 250–252
Haubrich, R., 336
Hausdorff, F., 310, 597
Hayashi, S., 83
head normal form, 779
Heck, R. G., 968, 991, 992
Hegel, G. F., 501
Heijenoort
  's correction, see *correction*254
  Jean van, 200, 201, 203–207, 214–218, 220, 221, 224, 227–230, 235, 245–247, 249–251, 253
Heijenoort, J. van, 174, 175, 602, 725
Heisenberg, W., 71
Hendricks, V., 348
Henkin
  completeness, 34
  model, 57
Henkin, L., 49, 312, 565, 586, 589, 737, 753
Henkin-style semantics, 489
HEO_n, 770
Herbrand
  's False Lemma, 208, 216–218, 220, see also *correction*254
    definition, 216
  's Fundamental Theorem, 203, 208, 213, 215–226, 237, 240
    definition, 220
    application, 232–233
  –Ribet Theorem, 238–239
  -style consistency proof, 237
  complexity, 214–215
    definition, 215
    example, 215
  disjunction, 214–222, 240–241
    definition, 215
    example, 215
  expansion, 208–209, 214–216, 221
    definition, 209
    example, 209
  Jacques, Sr. (Herbrand's father), 197–198, 250

universe, 208, 240–242
definition, 208
Herbrand's Completeness Theorem, 455
Herbrand's dissertation, 135
Herbrand's Theorem, 385, 393, 406, 407, 488
Herbrand's theorem, 147
Herbrand, J., 175, 196–253, 323, 360, 362, 366–369, 374, 379, 483, 517, 619, 624, 690, 714
Herbrand-Gödel recursive, 732
Herbrands's work, 141
hereditarily undecidable, 594
Hermite, C., 353
Hertz, H., 329
Hertz, P., 669, 691
Herzberger, H., 558, 959, 963, 971–973
Hesseling, D., 484
Hessenberg, G., 341, 888, 895, 896
heterological, 285
heterological paradox, 551
Heterologicality, 438, 439
Heyting algebra, 103
Heyting arithmetic, 667
Heyting's arithmetic, 93
Heyting's arithmetic of finite types, 94
Heyting's arithmetic of species, 115
Heyting, A., 77, 81–83, 85–89, 91–94, 96, 97, 99, 105, 110–115, 117, 203–204, 250, 375, 482–487, 491, 492, 495, 501–503, 525, 670–672, 677, 679, 683, 761–763
hierarchy
of implications, 942, 944
of universes, 939
hierarchy of languages, 544
higher-order logic, 173
higher-order metametalanguage, 44
Hilbert
's epsilon, 216, 218, 222, 224, 232, 233, 237, 252, 254
's finitism, see *finitism*254
's programme, 563, see *programme, Hilbert's*254
-style calculi, 223, see also *Hilbert, school, calculi of*254
school, 204–207, 212, 224, 227, 233, 237
calculi of, 212, 220
's 10th problem, 133, 175, 660
's finitist position, 168
's proof theory, 351
Hilbert, D., 79, 82, 87, 89, 90, 93, 94, 98, 104, 139, 172, 198, 200–202, 205–206, 212, 214, 216, 217, 220, 222, 227, 229, 231–233, 236–238, 245–247, 250, 251, 321, 350, 385, 387, 388, 390, 392, 393, 395, 402, 407, 410, 415, 418, 431, 434, 485, 490, 491, 495, 499, 501, 529, 530, 573, 583, 594, 607, 619, 621, 624, 625, 660, 671, 672, 698, 709, 714–716, 726, 728, 827, 879, 880, 882, 883, 885, 889, 895, 902, 903, 906, 907, 911, 918, 921, 926, 931, 932, 934, 945
Hilbert–Ackermann, 135, 172, 173
Hilbert–Bernays, 135, 172
Hilbert-style, 99
Hilbert-style formulation, 846
Hinata, S., 767
Hindley, J. R., 723, 739, 749, 773, 775, 776, 781, 840
Hinnion, R., 949, 993, 994
Hintikka, J., 994
Hobson, E., 887, 889
Hodes, H., 991
Hodges, A., 725
Hodges, W., 575, 608
Hoeven, G. F. van der, 83
Hofmann, K., 787
Holmes, M. R., 741, 993
homogeneity, 344

homomorphism, partial, 753
Honsell, F., 787, 977, 993
how-it-is, 35
Howard, W. A., 83, 684, 735, 759, 764, 767, 918
Hudelmaier, J., 701
Huet, G., 738, 772, 786
Hugh, F. C., 71
Hughes, G. E., 386, 417, 419, 428
Hume, D., 59, 60
Huntinton, E. V., 547
Hurwitz, A., 340
Husserl, E., 306, 309, 550, 752, 882, 895, 929
Hyland, M., 749, 771, 780, 784, 785, 788–791
hyperdoctrine, 757–758
hyperset, 978
hyperuniverse, 993
$Hyp(\mathcal{A})$, 944

ID_1, 982
ID_1^-, 940
ideal, 526
idealism, 257, 266
identity, 46, see also *tautology*254, 290, 292, 294
  axiom, 44
  conditions, 46, 62, 69
  normal, 214
    definition, 214
  predications, 44
Iemhoff, R., 83, 101, 102
IF, see *independence friendly first order language*875
ill-founded Tarskian hierarchy, 983
illative systems, 740–741
implected object, 62, 64
implection, 63, 68, 71, 72
implexive
  being, 59, 65, 69–72
  being and nonbeing, 65, 66
  nonbeing, 64, 69
  so-being, 64
implication, 41, 524
impossible, 32, 58
  constitutive property, 69
  Meinongian object, 71
  nonexistent objects, 64
  objects, 33, 37, 40, 48
impredicative definitions, 155, 156
impredicativity, 493, 495, 502
inaccessible cardinals, 600, 602, 603
incomplete, 32, 33, 64, 594
  beingless Meinongian object, 63
  beingless object, 62
  constitutive property, 69
  objects, 37, 40, 41, 58, 61
incompleteness, 667, 906, 916, 920, 983, 989, 990
incompleteness of formal systems, 618, 624, 627–633, 643–645, 655, 657
incompleteness theorem, 367, 369
incompleteness theorems, 462, 483, 491, 499
  first, 469, 498
  second, 470, 487, 499
inconsistent Meinongian objects, 58
independence, 114, 156, 514, 516
  of premiss rule, 101, 102
independence friendly first order language, 995
independent, 515
indeterminate object, 37
indifference thesis, 35
$Ind(\mathcal{A})$, 944
induction, 18, 392, 394, 407, 408, 410
inductive completeness, 934
inductive definability, 939, 943, 944, 997
inductive definition, 914, 933, 934, 937, 952, 966, 976, 982
  finitary, 924
  monotone, 950
  Uniform Monotone, 955
inductive set, 845
inference, 36

deep, 208, 213, 220, 224
principles, 39
rules of, see *Rule*254
structures, 39
infinite
the actual, 202, 204
infinitely proceeding sequence, 109–111
infinity
axiom of, 302
infintesimal analysis, 179
informal rigour, 501
information systems, 783, 787
inhabitant of a type, 776
Ben-Yelles' algorithm, 776
principal, 776
inherence, 69
inonsubsistent obect, 64
instability
degree of, 974
intended interpretation, 41
intended object, 29, 33, 59
intensional, 69
Meinongian semantics, 33
semantics, 34
set theory semantics, 46
intensional logic, 737
intensionality, 930
intensions, 46, 48
intentional
object, 35, 41, 42
intentionalist purpose, 35
intentionality, 29
of language, 34
of mind, 30
thesis, 30, 35
intermediate lgoic, 100
intermediate propositional logic, 102
internal determinacy, 34, 49
theorem for $\mathcal{O}$, 49, 50
internally determinate Meinongian object theory logic, 58
interpretability, 578, 579, 581
interpretation, 526, 532
intersection, 41
Introduction to Mathematical Philosophy, 2, 19
intuition of the continuum, 108, 109
intuition of time, 80
intuitionism, 200, 203–204, 235, see also *finitism* and *logic, intuitionistic*254, 324, 354, 853
intuitionistic analysis, 83, 108, 109
intuitionistic arithmetic, 77, 93
of arbitrary finite type, 116
intuitionistic first-order predicate logic, 77, 83, 89
intuitionistic logic, 387, 397, 399, 401, 670, 848
intuitionistic logic and arithmetic, 77
intuitionistic predicate logic, 96
intuitionistic propositional logic, 83–85, 91, 96
intuitionistic second-order logic, 77
intuitionistic sentential calculus, 525, 526
intuitionistic structure for the continuum, 109
intutionistic analysis, 114
invalid inference pattern, 49
invariance of dimension, 78
inversion principle, 765
invertible rules, 687, 694, 700
invertible terms, 746
involution, 967
Iongh, J. J. de, 84
Iongh, M. de, 84
IPC (intuitionistic propositional logic), 482, 483, 485–487
IQC (intuitionistic predicate logic), 482, 484, 487–489
irreducibility, 532
irreducible system, 526, 531, 534
Ishihara, H., 84
iterative classes of Post, 639, 640
Iverson, K. E., 744

Jäger, G., 936, 949, 955, 985, 996,

997
Jaśkowski, S., 309, 595
Jacobs, B., 764, 989
Jacquette, D., 30, 31, 33, 36, 46, 49, 58, 64–66
Jagadeesan, R., 785, 789, 790
Jané, I., 178
Janiszewski, Z., 598
Jaskowski, S., 103, 681, 700
Jeffrey, R., 404, 407
Jensen, R. B., 992
Jervell, H. R., 127, 179, 180
Jessop, T. E., 59
Jevons, W. S., 314, 340, 621, 633–634
Johansson, I., 682, 698
Johnstone, P., 787
join axiom, 955
Jongh, D. H. J. de, 84, 101–103
Jonsson, B., 584, 605, 608
Joyal, A., 790, 791
judgement, 277, 289
Jung, A., 783, 787, 788
jungle, 64, 65

(K), 863, *865*
König's lemma, 113, 116, 181
König's paradox, 536
König, J., 536, 878, 886, 887, 889, 890, 898
Kahle, R., 325, 882, 936
Kahn, G., 765, 785, 786, 791
Kaila, E., 698
Kalish, D., 411, 416
Kalman, J., 725, 776
Kalmar's class of elementary functions, 170
Kalmar, L., 711
Kamareddine, F., 725, 750, 997
Kanamori, A., 176, 882
Kanda, A., 788
Kant, I., 79, 81, 108, 201–202, 251, 287, 289, 301, 496, 501
Kaplan, D., 930, 980, 994
Kathail, V., 750
Kaye, D., 986
Kearns, J., 729
Keimel, K., 787
Keisler, H. J., 564, 565, 568, 602, 603
Kennedy, J., 83, 96, 98
Kerferstei, H., 333
Ketonen, O., 687, 698, 703
Keyser, C. J., 618, 620, 621
Kikuchi, M., 990, 991
Kindinger, R., 36
Kindt, W., 961–964
Kleene's number realizability, 104
Kleene, S. C., 42, 83, 88, 92–99, 104–106, 110, 113, 114, 374, 488, 494, 644, 650–998
Kleene's applicative structure, 770
Kleene-Rosser paradox, 732
Kline, G. L., 727
Klop, J. W., 738, 746, 749, 750
Kneale, W. C., 64, 401
Kneebone, G. T., 388
Kneser, H., 670, 677
Kochen, S., 565
Kœnig's Lemma, 227–229
Kohlenbach, V., 379
Kolmogorov's probelm interpretation, 88
Kolmogorov, A. N., 77, 83, 84, 88, 89, 96, 375, 484, 486, 670, 683, 990
Komori, Y., 750
Koons, R., 967
Korselt, E. W., 139
Korteweg, D. J., 78, 80
Kotarbiński, T., 306, 316, 319, 606
Kotlarski, H., 986
Kowalewski, J., 348
Koymans, C. P. J., 781, 786
Krajewski, Z., 986
Krajček, J., 952
Krauss, P., 97
Kreisel, G., 84, 108, 111, 114, 117, 237, 248, 249, 251, 252, 378, 391, 393, 394, 482, 485, 488,

492, 493, 500, 501, 704, 712, 744, 754, 757, 761, 762, 771, 780, 782, 900, 990
Kreisel, K., 706
Kripke model, 106–108, 487, 489
Kripke semantics, 103, 488
Kripke, S., 84, 106–108, 116, 117, 389, 399, 401, 938, 946, 961–967, 969, 971, 974, 976, 978, 983, 993, 998
Krivine, J. L., 723, 989
Krol', M. D., 84, 103
Kronecker, L., 166, 179, 204, 238, 324, 325, 327, 335, 340, 342–344, 352, 356, 378, 379
Kunen, K., 997
Kuratowski, C., 540
Kuzichev, A. S., 727, 741
Königsberg conference, 465, 472, 474, 486, 501, 502

Löwenheim
–Skolem Theorem, 130, 137, 173, 205, 215, 220, 224–229, 237, 564
à la [Löwenheim, 1915], 226
à la [Skolem, 1923b], 229
–Skolem–Tarski Theorem
Upward, 230
Löwenheim, L., 135, 139, 144, 172, 199, 205, 206, 210, 214, 216, 218–219, 224–229, 251, 453, 564, 567, 736
Läuchli, H., 84
Löb, M., 84, 956, 982
Löwe, B., 975
Lévy, L., 82
Laan, T., 725
labelled terms, 749, 750
Lachlan, A., 986
Lackey, D., 31
Lafont, Y., 789
Lamarche, F., 791
λ, 857
λ-calculus, 917, 919–921, 937, 939, 942, 945, 989
polymorphic, 989
lambda abstraction algebra, 781
lambda algebra, 781
λI-calculus, 860
λK-calculus, 860
lambda model, 781
lambda-abstraction, 35
$\lambda\beta$-calculus, 857
$\lambda\beta\eta$-calculus, 857
λ-calculus, 820, 856, *857*
$\lambda\eta$-calculus, 857
Lambek, J., 751, 752, 755, 758–989
Lambert, K., 36, 62, 70
Lamping, J., 750
Landin, P., 742–743, 747, 760, 777
constructors, 765
Landini, G., 891
Langford's result, 152
Langford, C. H., 486, 567, 568, 927
large cardinal axioms, 603
large cardinals, 598
lattice
κ-continuous, 942
non-modular, 968
lattice theory, 162, 172
lattices, 580
complete, 777
continuous, 777
Läuchli, H., 757, 762, 790
Lautman, Albert, 197, 248
law of contradiction, 17, 80
law of deduction, 528, 529
law of double negation, 77, 84
law of excluded middle, 22, 77, 80, 84, 113, 114, 517, 525
law of identity, 23
law of non-contradiction, 517
lawless sequences, 488, 495, 500, 501
Lawson, J., 787
Lawvere, F. W., 751, 755–759, 762, 771, 784
Lawvere, W., 988, 989

LCF project, 754, 784
Leśniewski, S., 305–319, 511, 529, 530, 550, 582, 595, 606, 752
Lebesgue, H., 888, 908
Leblanc, H., 49
Leeds, S., 556
Leibniz's Law, 64
Leibniz, G. W., 23, 24, 171, 494
Leibnizian identity conditions, 62, 66, 68
Leisenring, A. C., 386, 402, 411, 415
Leitgeb, H., 994
Leivant, D., 84, 769, 775
Lejewski, C., 318
Lercher, B., 736, 739
Lercher, Bruce, 839
Lescanne, P., 750
Lesniewski, S., 881, 902, 903, 911–913
levels of implication, 943, 944, 996
Levi, B., 886, 888, 895
Levitz, H., 176
Lévy, J. J., 749, 750, 781, 785
Levy, A., 605, 608
Lewis, C. I., 486, 619–621, 927
liar paradox, 423, 437, 438, 542, 543, 550
Libert, H., 994
Lifschitz, V., 84 972 972
limit rule, 973
limit-colimit coincidence, 778
Lindelöf, 353
Lindenbaum's lemma, 516, 518
Lindenbaum, A., 517, 518, 520, 521, 534, 594, 600
Lindenbaum, L., 583
Lindenbaum–Tarski algebra, 524, 525
linear lattices, 162
linear logic, 786, 789, 791
Lipschitz, R., 332
LISP, 742
lists, 769
locally confluent, 732
Loeb, I., 84
Loewen, Kenneth, 839
logic
  classical, 50, 199
  combinatory, 917, 918, 921–923, 937, 939, 942, 944, 946, 947
    Fitch's, 933, 934
  contraction-free, 955, 958, 995
  first-order, 205
  four-valued
    Dunn-Belnap's, 967
  illative combinatory, 937, 942, 945, 946
  intuitionistic, 203, 205, see also *intuitionism*254
  linear, 958, 996
  $\mathcal{M}$-, 986
  machine-oriented, 242
  many-valued, 955
  modal, 971, 980
  non-contractive, 923
  of fiction, 34
  ω-, 994
  partial, 949, 994
  substructural, 958
  three-valued, 924, 938, 952, 953, 964
    Łukasiewicz's, 970
    Kleene's, 960, 963
  two-valued, 199
logic of partial terms, 90
logic programming, 670
logical analysis, 21, 24
logical atomism, 21
logical axioms, 39, 40
logical consequence, 528, 530, 531, 536, 537, 559, 561, 562, 564, 565, 607
logical constants, meaning, 765
logical equivalence, 514
logical form, 21, 23
logical laws, 18
logical relations, 753
logical syntax, 43
logical truth, 561

logically equivalent, 515
logically ideal language, 23
logically inconsistent, 58
logically inconsistent Meinongian object, 58
logicism, 1, 12
Longley, J., 754
Longo, G., 781, 783, 787, 788, 989
Lorenz, K., 789
Lorenzen, P., 765, 789–791, 932–936, 939, 943, 966
Łoś, J., 565
Loux, M. J., 436
Luce, A. A., 59
Luckhardt, H., 84
ludics, 791
Łukasiewicz's 3-valued logic, 160
Łukasiewicz, J., 42, 305, 318, 518–522, 529, 776, 911–913, 955–957, 996
Lusin, N., 912
Lyndon, R., 565, 586

Mönnich, U., 997
Macbeth, 41
MacQueen, D., 774
Magari, R., 976, 988, 998
Majer, U., 324, 325
Mal'cev, A., 565, 788
Malacaria, P., 785, 789
Malitz, R. J., 949, 993
Mally, E., 29, 62
Mancosu, P., 83, 120, 176, 177, 325, 907, 926
Mann, C. R., 758
Mannoury, G., 84
many-valued system, 520, 521
Manzano, M., 725, 730
map theory, 946
Markov's principle, 105, 108, 114, 119
Markov's rule, 102
Markov, A. A., 84, 105
Martin, N. M., 634, 641
Martin, R. L., 558, 959, 961–963, 965
Martin-Löf, P., 84, 378, 684, 697, 748, 753, 769, 782, 787, 936, 939, 940, 989
  type theory, 763–765
Mason, J., 997
Mass terms, 386, 413, 414, 425
Massachusetts Institute of Technology, 825
material implication, 84
material spatiotemporal object, 71
mathematical epistemology, 18
mathematical identity, 90
mathematical induction, 14, 56
mathematical logic, 3
mathematical notation, 38
mathematics
  finitist, 368
Matiyasevich, Y., 176
matrices, 522
matrix, 519, 520
matrix method, 519–522
Mauny, M., 786
maximal consistent subset, 981
maximally
  consistent, 52, 55
  impossible beingless object, 70
  impossible object, 65–69, 71
Mazurkiewicz, S., 598
McCall, S., 42
McCarthy, J., 742, 744, 754
McCarty, D. C., 84, 488
McCracken, N., 783
McCusker, G., 791
McGee, V., 967, 968, 972, 974, 975, 983, 994
McGuinness, B., 275
McKinsey, J. C. C., 486, 572
McNulty, G., 580, 608
McTaggart, J. M. E., 3
measurable cardinals, 603
Medvedev, Ju. T., 762
Megill, N., 776
Mehra, J., 71
Meinong's

concept, 66
implection theory, 69
Jungle, 65
object theory, 64, 72
object theory semantics, 72
theory, 65
Meinong, A., 29–34, 36, 41, 59–62, 64, 65, 69. 70, 71
Meinongian, 34, 35
definite descriptor, 68
desccription and abstraction, 68
domain, 46, 68
extraontology, 66
implection, 71
logic, 29–58
object theory, 33, 46
objects, 41, 62, 67
One, 71
Permenidean One, 72
predicate object theory O, 57
quantification, 68
quantificational logic, 41
semantic domain, 41, 59, 69
semantic domain's extraontology, 64
set identity axiom, 45
set membership, 43
set theory, 43, 45, 46
theory of defininite description, 67
memory, 30
Mendler, N., 997
menge, 109
Menger, K., 483, 485, 498, 671, 736
mental acts, 33
Menzler-Trott, E., 668, 698, 706, 709, 714, 715
Meredith, C. A., 759, 776
Meredith, D., 776
mereological physical containment, 70
mereologically, 61
mereology, 306, 307, 311, 316
metalanguage, 544, 545, 549–554, 559
metalanguages, 44
metalogic, 308, 512, 529
metamathematics, 79, 84, 512, 526, 529
metaphysical systems, 72
metaphysically impossible, 58
metaphysics, 72
metasystem, 537, 538
metatheoretical device, 40
metatheory, 69, 549, 552
metatheory of first order arithmetic, 179
method
axiomatic, 326, 334, 351
genetic, 326, 334
method of eliminating quantifiers in decision problems, 151
method of elimination of quantifiers, 564, 567, 581
method of proof by interpretation, 528
method of quantifier elimination, 568, 575
Meyer Viol, W. M. P., 399, 404, 405, 410, 411
Meyer, A. R., 781
Meyer, F., 79
Meyer, R. K., 417, 420–422, 424, 428, 776, 781
Mezghiche, M., 739
Mill, J. S., 306, 389
Milner, R., 754, 784
ML, 754, 775
typability algorithm W, 775
Minari, P., 955, 968
minimal complete development, 748
minimal logic, 682, 848
minimalism, 968
Mints, G., 84, 101, 176, 392, 398, 776
Mirimanoff, D., 153, 764, 903–906, 928, 988, 989
Mislove, M., 787
Mitchell, J. C., 774
Mitschke, G., 749, 750
ML language, 754, 775, 786
model, 526–528, 561, 592

N_k-, 993
ω-, 982, 983
recursively saturated, 986
model consequence, 588
model theory, 526, 536, 564, 565
models of typed λ, 737, 753, 770–771
general and standard, 737, 753
of polymorphic λ, 770, 771, 786, 789
PERs, 771
types in $\mathcal{P}(\omega)$, 782
models of untyped λ, 773, 777–789
approximation theorem, 779
as inverse limits, 778
continuous lattices, 777
D_∞ (Scott), 777–778, 783, 784
filter models, 773, 787
graph models, 781–782
Engeler's model, 782
$\mathcal{P}(\omega)$, 782–783
$\mathbb{T}^\omega$ (Plotkin), 783
reflexive domains, 778
several definitions, 781
universal domains, 783
modern logic, 6, 24
modular lattices, 580
modus ponendo ponens, 40
modus ponens, 80, 203, 212, 218, 220–224, 238, 826
definition, 212
elimination, 203, 218, 221–224, 238
Moerdijk, I., 84
Moggi, E., 771, 788, 789
Monk, D., 586, 589, 608
monoid structure of CL, 746
monotransformable set of sentences, 536
Montague grammar, 753
Montague, R., 411, 416, 423, 438, 565, 930, 980, 994
Moore, G. E., 1, 255, 275, 287, 597, 600, 608
Moore, G. H., 325, 878, 880, 881
Morel, A. C., 565
Morgan, A. de, 583
Morley, M., 652–653
Morris, J. H., 747, 766, 772, 774–787
typability algorithm, 775
Morse, M., 501
Moschovakis, J. R., 97, 114
Moschovakis, Y., 120, 935, 944, 964, 976, 997
Moser, Georg Ch., 244
Moses, J., 742
Moss, L., 979, 997
Mostowski, A., 103, 565, 575, 577, 606, 607
Mow-Shaw-Kwei, 956
MP (Markov's Principle), 482, 487–489, 495, 502
Mt. Analogue, 37, 41, 42
(μ), *858*, 865
Muchnik, A. A., 648, 654
multiple-relation theory of judgement, 2
Muller, F. A., 178
multiplicity, 879, 880
Murthy, C., 766
My Philosophical Development, 2
Mycielski, J., 597
Myhill, J., 84, 378, 746, 771, 782, 920, 922, 924, 928, 933–937, 939, 942–946, 961, 966, 996, 998

n, *860*
Nagel, E., 330
naive comprehension principle, 46
naive set theory, 7
Nakajima, R., 745, 780
names, 260
natural deduction, 99, 667, 672, 695, 829, 846
natural languages, 544
NBG set theory, 727
Neale, S., 432
Nederpelt, R. P., 725, 739, 750

negation, 41, 42, 77, 674
external, 926
internal, 925, 978
negation as absurdity, 84
negative translation, 97, 484
Negri, M., 977
Negri, S., 694, 697, 702
Nelson's theorem, 104
Nelson, D., 84, 88, 104–106
Nelson, L., 878, 888, 895–898, 995
Nemeti, I., 589
neo-Meinongian
extraontology, 68–70
logic, 68
ontology, 69
semantic domain, 69
semantics, 68
Nepeivoda, N., 961, 966
Nerode, A., 652–653, 777
Neumann, J. von, 152, 198, 206, 229, 233, 235, 251. 727
New Foundations, see *NF*875
new realism, 5
Newell, A., 24
Newman, M. H., 732, 734, 738
typability algorithm, 775
Newton, I., 24
next stable structure, 976
NF, 930, 992, 993
NF ("New Foundations"), 734, 741
NFU, 992, 993
Nickau, H., 791
Nidditch, P. H., 59
no-classes theory, 878, 890, 891, 897
Noether, Emmy, 198, 238, 251
nominalism, 60, 64, 316, 318, 319, 606
nominalization, 996
Non-attributive reference, 389, 416, 417, 436, 443
non-Euclidean geometry, 593
non-extensionality, 916, 946, 947, 949, 951, 954
non-Fregean, 46
non-standard models of arithmetic, 474
non-standard models of axiomatic number theory and set theory, 134
nonbeing, 35, 59, 63, 69, 71, 72
nonentity, 41
nonexistent, 31, 71, 72
nonsubsistent object, 64, 65
objects, 34, 37, 41, 42, 48
nonlogical axioms, 40
nonsense, 288, 299
nonstandard analysis, 179
nonstandard models for arithmetic, 156, 158, 178
nonstanrdard set theory semantics, 34
nonstardard deduction theorem, 34
nonsubsistent object, 66
normal form, 732, *859*
approximate, 779
head, 779
normal form theorem, 837
normal inhabitant, 776
normal matrix, 521
normal systems of Post, 623, 627–633, 644, 648, 651, 656
normalization, 678, 766
proofs, 735, 737, 740, 747, 766–767, 772, 773
notation
uniform, 210–211
Novikoff, P. S., 791
normal form, 130
(ν), *858*, 865
nuclear, 58
(constitutive) property, 34
-extranuclear property distinction, 36
-unrestricted comprehension, 45
constitutive property, 46
predications, 47, 48
property, 36, 37, 58, 63
relational properties, 43
number

cardinal, 878, 879, 881, 885
ordinal, 878–880, 885, 896
numerals
of Böhm, 746
of Scott and others, 744
of Church, 731, 769
numeralwise expressibility, 468
Nuprl system, 765

ob, 845
ob-system, 845
object, 29, 30, 44, 256, 257, 259, 261, 262, 264, 271, 292
constant, 38
form of, 266
language, 46, 544, 545, 549–551, 553, 554, 559
theory, 34, 65, 549
theory $\mathcal{O}$, 37, 39
theory (*Gegenstandstheorie*), 29
theory semantic domain, 58
theory semantics, 41
variable, 38
with being, 32
objectives, 32
observationally equivalent, 784
Ockham, W., 313, 891, 903
Odifreddi, P., 976, 990
O'Donnell, M. J., 738
Ogasawara, T., 103
Oksanen, M., 993
ω-consistency, 468
Omega Rule, 395, 410
ω-consistency, 824
ω-inconsistency, 983, 996
ω-rule, 916, 932
ON, 879, 887, 890, 896, 905
One, 71
1, *860*
Ong, L., 785, 789–791
Ono, K., 698
ontic
domain commitment, 69
neutrality, 72
status, 71
ontological
commitment, 34
import, 41
status, 35
ontology, 33, 64, 307, 314
ontology of existent or subsistent objects, 64
Oosten, J. van, 84, 106, 120
van Oostrom, V., 738
operational semantics, 765
operative approach of Lorenzen, 180
operator
abstraction, 918, 947, 948, 952
equivalence intensional, 952
intensional equivalence, 952
jump, 964, 972, 983
logical, 933, 938, 947
modal, 896, 994
monotone, 923, 937, 941, 978, 979, 998
monotone three valued, 938
positive elementary, 939, 940
order of a function, 899
order of a language, 549
ordered n-tuple, 48
ordinal
closure, 964
reflecting, 974
stabilization, 973
ordinal addition, 604
ordinal algebras, 605
ordinal analysis, 933
ordinal number, see *number, ordinal*875
ordinal-definable sets, 658–659
Ore, Ø., 129
Orilia, F., 997
Orlov, I. E., 486

$\mathcal{P}(\omega)$, *see* models of untyped λ
Péter, R., 168
PA, see *Peano Arithmetic*875
PA (Peano Arithmetic), 460, 467, 468, 474–476, 481–485, 490, 491,

493
Padoa's method, 534
Padoa, A., 531, 532, 534
pairing, 45
de Paiva, V., 789
Pani, G., 791
paradox, see *Russell's Paradox* and *Skolem's Paradox*254, 338, 536, 538, 539, 553, 554
Banach-Tarski, 877
Berry's, 888, 892, 898, 907, 989, 990
Burali-Forti's, 877, 879, 881, 887, 889, 890, 893, 896–898, 904, 905, 930, 989
Cantor's, 879–881, 893, 897
Curry's, 917, 923, 942, 956
Girard's, 989
Grelling's, 897, 926, 929
Grelling-Nelson's, 888
Hangman, 980
hypergame, 988
König's, 886, 890, 898
Knower's, 980
Liar, 876, 891–893, 895, 898, 902, 908, 912, 913, 916, 928, 934, 937, 959, 965, 966, 975, 977, 979, 994–996
Mirimanoff's, 988, 989
Newcomb's, 877
of Curry, 733
of Girard, 764
of Kleene and Rosser, 732
of vagueness, 877
Richard's, 892–894, 896, 898, 900, 906–908, 910–912, 916, 917, 920
Russell's, 877, 881, 882, 890, 893, 894, 896–898, 906, 908, 918, 919, 922, 923, 926, 928, 934, 936, 937, 940–942, 968, 991, 993
semantical, 876–878, 893, 910–912, 916, 956, 959, 968, 971, 977, 978
Yablo's, 987, 988
parallel computation, 784
parallel conditional, 784
parallel or, 785, 786
parallel reductions, 748
parametricity, 769
Paris, J., 958, 996
Paris-Harrington Theorem, 476
Park, D., 778
Parmenidean idealism, 72
Parmenideanism, 71
Parmenides, 71, 72
Parsons, C., 41, 374, 497, 502, 959, 979
Parsons, T., 36, 37, 325
part-whole relations, 61
partial homomorphism, 753
partiality, 919, 929, 933, 997
Pasch's axiom, 590
Pasch, M., 530
Paulson, Lawrence C., 195, 244
Pavlovic, D., 989
PCF language, 755, 784–786, 789, 791
Peano, see also *axioms, Dedekind–Peano*254
Peano Arithmetic (PA), 575, 576, 667, 939, 940, 966, 968, 971, 980, 982, 983, 985–987, 990, 992, 994, 996
Peano, G., 1, 2, 5, 6, 8, 24, 79, 81, 87, 177, 200, 205, 206, 313, 499, 530, 531, 725, 731, 877, 878, 881, 883, 888, 889, 891–894, 944
Peckhaus, V., 324, 338, 882
Pegasus, 30, 32, 71
Peirce
–Schröder tradition, 205, 216, 231
–Schrœder tradition, 227
Charles S., 200, 205, 208, 209, 216, 247, 252
Peirce, C. S., 317, 583, 736
Pelham, J., 891

Pelletier, F. J., 634, 641
PER model, 771
perception, 30
Perlis, D., 997
Personal pronouns, 385, 418, 432, 433
Petersen, U., 995
phenomenal world, 71, 72
phenomenlogical
tradition, 29
phenomenological, 32, 60
standpoint, 58
phenomenology, 501
Phlogiston, 30–32
Phoa, W., 789
physical reality, 71
Π (restricted implication in Church's system), 919
picture theory, 261
Pierce, B. C., 774
Pieri, M., 594, 901
Pigozzi, D., 522, 524, 526, 781
Πm, 884
Pinsent, D., 280
Piperno, A., 729
ΠR, 884
Pitts, A., 771
Platek, R., 754, 976
Plato, J. von, 99, 173, 174
Platonic subsistence, 35
platonism, 853
Plotkin, G. D., 765, 770, 774, 778, 780, 782–788, 791, 937, 946, 997
$\mathcal{P}(\omega)$ model, 781
$\mathbb{T}^\omega$ model, 783
pluralism, 5
Pogorzelski, W., 607
Pohlers, W., 378
Poincaré, H., 10, 80, 339, 346, 348, 353, 204, 886, 888–893, 895–898, 900, 902, 904, 908, 911, 912, 918
polarity, 977
polyadic algebras, 736
polymorphism in types, 734, 771
models of, 770, 771
parametric, 768
Popper, K., 700
Possel, R. de, 572
possible-world semantics, 106
Post
-completeness, 622, 659
-consistency, 622
correspondence problem, 648, 649, 659
machines, 641–643
productions, 626–627
Post's
Principia Mathematica, 618–622, 625, 627–628, 631–633
lattice, 634–639
problem, 643–648, 654
theorem, 651–652
thesis, 628–629, 631–632
Post, E., 175, 515, 516, 714
Posterior Analytics, 529
Pottinger, G., 766, 773
Poutsma, H., 501
power class-existence
strong, 955
weak, 955
power set, 45
axiom, 46
power set axiom, 600
Powers, L., 781
Prawitz, D., 678, 682, 696, 764, 765, 768, 995
Pre-suppositions, 387, 388, 418, 434
predicate, 38, 46
complementation, 38, 40
expression, 42, 44
object theory logic, 54
partially defined, 959, 960, 964
semantics, 46
symbol, 37
total, 945, 969
trial and error, 976
predicate logic, 586, 588, 589

predication, 34, 40, 313
predicational incompleteness, 71
predicative function, 890, 900, 908, 909
predicative functions, 11
predicativity, 888, 890, 895, 899, 900, 927, 983, 984
prenex, see *direction, prenex* and *form, prenex*254
Presburger
  Arithmetic, 231, 241
  Mojżesz, 231, 252–254
Presburger, M., 93, 96, 362, 567, 575
Priest, G. G., 388, 442, 443, 968
Priestly, J., 30, 31
primary substance, 60, 72
primitive operation, 845
primitive recursive arithmetic, 94, 357
primitive semantic relation, 41
primitive syntax, 38
Princeton University, 826
principal functional character, 841
principal type, 774
  algorithms, 774–776
principal type scheme, 841
Principia Mathematica, 2, 3, 10, 12, 13, 15, 18, 24
Principia Mathematica, 517, 545
Principia Mathematica, 200, 206, 212, 220, 254
  calculi of, 212, 220
principle of bar induction, 113
Principle of Dependent Choice, 227
Principles of Mathematics, 1, 9, 10, 12, 22
Prior, A. N., 436
priority method, 648, 654
problem of finite axiomatizability, 592
problem of universals, 59
process
  almost everywhere periodic, 973
  B-, 974, 976
  Γ-, 975
  G-, 975, 976
  H-, 972–974, 976
  non terminating, 926, 927
  P-, 974, 975
  revision, 973, 975
program composition, 835
programme
  Fifth Generation Computer, 243
  Hilbert's, 200–202, 205–206, 233, 236–238
  unwinding, 237–238
projective geometry, 593
proof, 566
  intuitionistic notion of, 491–495
Proof Interpretation, 482, 487, 488, 491–495, 502
proof search, 223–224, 240–242
  automatic, 223, 240–242
  human-oriented, 223
proof theory
  Hilbert's, 323
proof-net, 790
proof-theoretic strength, 940, 980
proofs as objects, 761
proofs as terms, see *deductions as ...*735, 759
proper implection, 68, 70, 71
proper subset, 62
proper subset membership, 63
properties, 32, 33, 37, 66
Property A, 206, 214, 218, 220, 224
  definition, 214
Property B, 214
  definition, 214
Property C, 214–221, 224–227, 230, 232, 237, 240
  definition, 215
  example, 215
property implection, 66
property theory, 876, 930, 993, 997
propositional and predicat einference, 34
propositional calculus, 620–623, 633–641
  completeness, 620–622

decidability, 620–622
functional completeness, 622, 634, 640–641
generalized, 622–625
undecidable, 650
propositional connectives, 88
propositional function, 883, 884, 889, 890, 894, 898–900, 907, 909, 919, 923, 926, 929, 940, 941
propositional functions, 6, 7, 11, 13, 24
propositional logic of intuitionistic arithmetic, 102
propositional negation, 38, 40
propositional object theory logic, 50
Propositional Operators, 410, 421, 422, 424, 426, 436, 437
propositional variables, 38, 40
propositions, 32, 66
propositions as types, 734–735, 756, 759–766
in classical logic, 766
propositions-as-types, 838
prothothetic, 307, 311, 312
provability, 554, 567, 573
provability logic, 487
provability operator, 486, 487
pseudo-proposition, 300
psychological acts, 29
psychologism, 29
psychologistically, 30
Pudlak, P., 179
pure λ and CL, 723
pure λ-calculus, 823
Pure Type Systems, 741

qualities, 32
quantifier
δ-, 211
γ-, 211
general, 211
restricted, 211
quantifier elimination, 491
quantifier elimination theorem, 568, 570–572
quantifiers, 6, 7, 89, 90, 318
quasi-leftmost reduction, 749
Quine, W. V. O., 31, 41, 59, 65, 317, 319, 416, 423, 425, 556, 607, 730, 734, 736, 875, 890, 924, 929, 930, 969, 980, 992, 993
system NF, 734, 741

$\mathrm{RA}_{<}\alpha$, 984
van Raamsdonk, F., 738
Rabin, M., 565
Raggio, A., 682
raising, 229
Ramsey's theorem, 147, 149
Ramsey, F. P., 270, 284–287, 293–297, 737, 751, 877, 903, 907–909, 911, 929, 992
Range theorem, 746
Rapaport, W. J., 65
Rasiowa, H., 84, 99, 103
Ratajczyk, Z., 986
Rathjen, M., 713, 955
Ravaglia, M., 325
real closed field, 592
real number generator, 111, 112, 114, 115
realisation, 526
realism, 14, 264–266, 279
realizability, 494
realizability semantics, 103
realizability, modified, 761
realizability, recursive, 762
realizability-$\vdash$, 105
realizes, 104
recursion induction, 742, 754
recursion theory, 633, 641–654
recursive arithmetic, 131, 134, 166
recursive definition, 363
recursive function theory, 169
recursive functions, 732
recursive method, 547, 556
recursive number theory, 364

recursive realizability, 103, 104
recursive types, 774
recursively enumerable sets, 629, 643–648, 651–652
  creative, 645–647
  hyper-simple, 647
  simple, 646, 647
recursiveness conditions, 374
redex, 732, 738, *859*
  β, *859*
  η, *859*
  K, 863, *865*
  S, 863, *865*
  weak, 865
reducibility, 624, 645–648
  many-one, 645–646
  one-one, 645
  truth-table, 646–647
  Turing, 645, 647–648, 651–654
reducibility axiom, 890, 907, 908, 927
reduction, 857, *858*
  $\lambda\beta$-, 732, 748–750
  $\lambda\beta\eta$-, 739, 749
  abstract, 738
  β, *858*, 859
  $\rhd_\beta$, 859
  $\beta\eta$, *859*
  computational, 378
  efficient, optimal, 750
  foundational, 378
  inside-out, 749
  parallel, 748
  quasi-leftmost, 749
  Révész', 750
  $\rhd$, *858*
  strong, 838
  strong in CL, 739
  weak, 863, *865*
  $\rhd_w$, 863, *865*
    for λ-terms, *866*
  weak in CL, 739, 748
reduction classes, 146
reduction diagrams, 749
reduction problems, 134
reduction strategies, 749
reference, 66
Ref(PA), 984, 985
Ref*(PA(P)), 984
reflexive domains, 777, 778
refutability, 848
regressive method, 16, 17, 20
regular cardinals, 602
regular ordinal, 601
Reid, C., 709
relation algebras, 580, 583, 598, 606, 736
relation arithmetic, 13
relation numbers, 15
relational algebra, 589
relational algebras, 584
relations, 32
relative interpretability, 580
relatively interprtable, 580
relativised Hilbert Programs, 475
relevant implication, 737
relizability, 114
replacement, 46
representability condition, 369, 370
representation problem, 592, 594
Rep($\mathcal{A}$), 944
Rescher, N., 42
residuals, 738, 749
Resnik, M., 556, 558, 608
Restall, G., 995, 996
restricted generality, 834, 841
restricted the axiom of comprehensionto arithmetical properties, 181
restriction, 46
retraction in $\mathcal{P}(\omega)$, 783
Révész, G., 750
revision procedure, 971
revision process, see *process, revision*875
revision theory, 974, 975, 997
Reynolds, J. C., 768–770, 774, 779, 784, 989
Rezus, A., 783
(ρ), *858*, 865

Richard's Paradox, 442, 536, 823, 832
Richard, J., 353, 536
Richman, F., 84
von Rimscha, M., 781
rings, 580
    commutative, 580
    ordered, 580
Rios, A., 750
Robinson, A., 179, 565, 575, 985
Robinson, J. A., 170, 240, 242, 252, 565, 580, 775
Robinson, R. M., 577, 581
Rogers, H., 782
Rogers, H. Jr., 935
Ronchi della Rocca, S., 774
Roquette, Peter, 197, 198, 203, 207, 244, 245, 251, 252
Rosen, B. K., 738, 750
Rosenbloom, P., 623, 659, 729, 733, 737
Rosolini, G., 788
Ross, J., 93
Rosser, J. B., 95, 98, 575, 714, 725, 729, 731, 749, 822, 910, 917, 920–924, 929, 930, 953, 970
round square, 30, 33, 37, 43, 65, 66
Routley, R., 36, 64, 417, 420–422, 424, 428
Rouyer-Degli, J., 750
Rowe, D. E., 325
Rubin, H., 600
Rubin, J., 600
Russellian definite description theory, 67
Ruitenberg, W., 84
Rule, *see also* Generalized Rule
    of δ-Quantification, 220
        definition, 214
    of γ-Quantification, 220, 223
        definition, 214
    of Generalization, 214
    of Implication, 214
    of Passage, 213, 214, 216–218, 220, 224
        definition, 213
    of Simplification, 214, 218, 220, 222
        definition, 214
rule, 845
rule of inference, 94
Russell, see also *Principia Mathematica*254
Russell's paradox, 1, 7, 8, 10, 12, 305, 306, 310, 314, 660
Russell, B., 1–25, 31, 65, 79, 87, 89, 135, 199, 200, 202, 205, 206, 212, 216, 220, 255–260, 264, 267, 269, 275–277, 282, 285–287, 289–291, 302, 307, 309, 313, 314, 317, 323, 341, 342, 346, 359, 385, 386, 387, 389, 416–418, 430, 435, 439, 499, 536, 550, 600, 618–622, 625, 672, 725, 728, 731, 735, 751, 764, 769, 821, 822, 877, 878, 880, 881, 883–885, 887–900, 904, 907, 918, 922, 926–928, 978, 988
Russian recursive mathematics, 105

(S), 863, *865*
S4, 482, 485–487
Słupecki, J., 318
Saarinen, E., 435
Sacks, G., 652–653
Salibra, A., 781
Salkowski, E., 177
Sallé, P., 773
Sambin, G., 788
Sanchis, L., 767, 840
Sandu, G., 994
Santa Claus, 32
satisfaction, 539, 540, 547–549, 561
satisfaction class, 986
satisfaction of a sentential function, 536, 537
Sattler-Klein, Andrea, 244
saturation, 117, 118

saturation point, 973
Sayward, C., 438
Sazonov, V. Y., 784
Scanlon, Thomas M., Jr., 210, 233, 237, 249, 252
Scarpellini, B., 84
Scedrov, A., 84
Schönfinkel, M., 726–728, 730, 826, 917, 918
Schütte, K.,84, 98, 378, 699, 712, 768, 931–933, 963, 996
Schönflies, A., 888, 892–896, 900
Schütte, H., 98
schema
  choice, 945
  comprehension, 926, 929, 931, 941, 948, 949, 951, 958, 968
  T-, 912–915, 938, 968, 969, 975, 996
Scheme (programming language), 742
Schmidt, A., 685, 687, 714, 715
Schoenflies, A., 110
Scholastic distinction, 32
Scholz, H., 714
Schröder, see also *Peirce-Schröder tradition*254
Schröder, E., 200, 205, 210, 216, 252, 313, 314, 583, 619, 622, 736
Schröder, M., 135, 139
Schroeder-Heister, P., 670, 697
Schroer, D. E., 738, 739, 749
Schwichtenberg, H., 378, 701
scientific outlook, 5
scientific thought, 64
Scott, D. S., 84, 103, 179, 565, 602, 744–745, 757, 760, 779, 782–784, 786–788, 922, 937, 939, 940, 942, 943, 945, 961
  continuous lattices, 777
  first model of λ (D_∞), 777–778
  LCF system, 754
  numerals, 744
  reflexive domains, 778
    as inverse limits, 778
  Scott domains, 786–787
  Scott topology, 777
  synthetic domain theory, 788–789
  theory of constructions, 761–763
  undecidability theorem, 744
Scott-continuous function, 777
Scott-open set, 777
Scowcroft, P., 84
SECD-machine, 743, 744
second act of intuitionism, 77, 108
secondary substance, 60
Seely, R. A. G., 790
Segerberg, K., 926
Sein -indifferent, 62
Sein-independent, 62
Selby-Bigge, L. A., 59
Seldin, J. P., 723, 725, 728–730, 739, 740, 742, 771, 772, 840, 917
self, 278, 279
self-application, 730, 882
self-reference, 876, 877, 889, 903, 910, 928, 963, 977, 987, 988, 998
  direct, 882, 917, 922, 924, 977, 978
  indirect, 910, 922
Self-truth axioms, 984
Seligman, J., 997
semantic category, 550
semantic conception of truth, 558
semantic definability, 531, 537
semantic paradox, 544
semantical category, 549
semantical paradoxes, 559
semantically closed, 542, 543, 937, 964, 978, 998
semantically universal, 543, 544
semantics, 31, 41, 72, 536
  consistency, 54
  domain, 41
  inconsistency, 54
  model, 41
  of predicate logic, 44
  theory, 29
semi-inductive process, 971–974, 997

semi-separability of λ-terms, 780
semi-Thue system, 627, 649–650
semigroups, word problem, 648–650
semtically universal, 542
sense, 256, 257, 267, 268
sentence
  grounded, 963
  Liar, 892, 899, 972, 978, 979
  paradoxical, 960, 965, 971, 974
  ungrounded, 965
sentential calculus, 513, 516, 518, 520, 524, 525, 547
sentential function, 532, 537–540, 546–549, 551, 561
sentential tautology, see *tautology, sentential*254
separability of λ-terms, 780
separated system, 841
sequent calculus, 667, 677, 684, 691, 695, 698
sequentiality, 785–786
  sequential algorithm, 785
  sequential function, 785
    on concrete domains, 785
series, 13
set
  (completely) representable, 944
  (formally) discrete, 945
  admissible, 976, 994
  circular, 905, 906
  Δ^1_1, 966
  Δ^0_2, 976
  denumerable, 894
  discrete, 946
  effectively enumerable, 894
  extraordinary, 904
  grounded, 904
  hyperarithmetical, 934, 935, 966
  internally definable (in Frege structures), 941
  logical notion of, 940–942
  non-well-founded, 904, 905, 977
  of first kind, 904
  of grounded set, see *WF*875
  of second kind, 904
  ordinary, 904
  partial, 962, 964
  Π^1_1, 935, 966
  Π^1_2, 967, 975, 976
  positive, 993
  positive inductive, 935
  separable, 946
  Σ^1_1, 967
  strongly definable, 976
  transitive, 993
  unbounded, 974
  weakly definable, 975
  well-ordered, 879, 886
set theory, 580, 598, 606, 876, 877, 879, 880, 888, 889, 893, 895, 896, 901, 903, 905–907, 911, 928, 930, 988, 993
  Cantorian, 878, 879
set theory, finitely axiomatized, 727
set-theoretic relativism, 133
Sethi, R., 774
sets, 66
  abstraction, 46
  identity axiom, 44
  membership, 44
  theoretical semantics, 46
  theory, 44, 46
  theory paradoxes, 46
Shaw-Kwei, M., 171
Sheard, M., 980–982
Shepherdson, J. C., 160, 771, 782, 958, 996
Sherlock Holmes, 32
Shirahata, M., 996
Shoenfield, J. R., 58
Sieg, W., 725, 730, 885
Siekmann, Jörg, 195, 242, 248, 249, 251, 252
Sierpiński, W., 310, 311, 598, 600
(σ), *859*, 865
Σ (existential quantifier in Church's system), 919
Σ_∞, 942, 944, 945

sign, 273–275
significist, 84
Sikorski, R., 103
similarity, 13–15
Simmons, K., 551, 558, 902, 967, 979
Simon, H., 24
Simons, P. M., 70
simple theory of types, 531, 534, 537, 562
simple type theory, see *types, simple*737
Simpson, S., 378
situation theory, 977, 979, 997
Skokem, T., 130, 205, 206, 210, 214, 216, 219, 226–231, 241, 252–254, 410, 453, 564, 567, 699, 703, 901, 956–958, 995, 996
Skolem, see also *form, Skolem normal* and *form, Skolemized* and *function, Skolem* and *Lœwenheim–Skolem Theorem* and *Lœwenheim–Skolem–Tarksi Theorem* and *Skolemization* and *term, Skolem*254
's Paradox, 229
Skolem 1920, 453
Skolem 1922, 454
Skolem 1928, 454
Skolem functions, 130, 134
Skolem Hull construction, 178
Skolem normal form theorem, 136
Skolem's paradox, 127, 155, 177
Skolem–Noether, 132
Skolemization, 210
Skyrms, B., 959
Slomson, A. B., 179
Smielew, W., 580
Smith, James T., 244, 251, 252
Smith, S., 986
Smorinski, C., 84, 910, 956
Smullyan
's uniform notation, see *notation, uniform*254
Raymond M., 211, 253
Smullyan, R., 969
Smyth, M., 784
Smyth, M. B., 778, 784, 787, 788
SN, 766, 767, *see* normalization
so-being, 33, 35, 43, 59, 64, 69
Soames, S., 558, 559
Sobociński, B., 314, 318
Sobocinski, B., 518, 520
solid geometry, 595, 596
solipsism, 266
Solving solution, 391
Sosein, 35, 37, 41, 58, 64, 66, 67, 70, 71
function, 63, 67, 68
set, 63
Sosein set theoretical interpretation, 72
Soseine, 43, 46, 48, 59, 62, 63
soundness, 103
space, 595, 887, 945, 960, 962, 964
injective T_0-, 945
compact Hausdorff, 988
discrete, 945
exponential, 993
function, 921, 951
κ-, 945
of truth-values, 967, 969
spatio-temporal entities, 65, 69
spatio-temporal existence, 35
spead, 108
species, 109, 115
species of first order, 109
species of second order, 109
Specker, E., 903, 905, 992
Spector, C., 83
spheres, 595, 597
spread, 109, 110, 117
spread law, 110, 117
squadge, 938, 944
Stärk, R., 997
stable element
negatively, 972
positively, 972
stable function, 786

stable truth, 973
Stalnaker, R. C., 416, 417, 420, 422, 428
standard models, 737, 753
Standardization Theorem, 839
standardization theorem, 749
Staples, J., 750
state of affairs, 256, 259–262, 289, 290, 977–979
Statman, R., 753
Steele, G. L., 742
Steen, S. W. P. , 405
Steinby, M, 738
Steinitz, W., 174
Stenlund, S., 723
Stigt, W. P. van, 81, 84, 110
Stone, M. H., 103, 525, 526, 586
Stoy, J., 777
Strachey, C., 768, 777, 779
stratified formula, 734, 774
Streicher, T., 789
strict refutability, 848
strong completeness theorem, 54, 57
strong counterexample, 113, 114
strong equivalence, 45
strong normalization, see *SN*766
strongly inaccessible cardinals, 601
strongly inacessible, 602
structural rules, 685, 696, 701
structural-descriptive names, 545, 547, 549, 551, 552, 554
structural-descriptive terms, 553
structuralism, 854
structured operational semantics, 765
structured set (menge), 109
Studnet, W., 715
Sturm's theorem, 571
Størmer, C., 179
subsets, 45
subsistence, 35, 69, 72
subsistent
  object, 62, 70, 71
subspecies, 115
substance, 257
substitution, 826, *858*, 863, 864
  Curry's analysis, 728
  efficient, 750
  explicit, 750
  $-[-/-]$, *858*
  very early notations, 725
substitutional theory, 910, 967
Sullivan, P., 296
Sundholm, G., 765
supervaluation, 937, 960, 961, 965, 966, 982, 997
Suppes, P., 330, 608
Surma, S., 607
symbol, 273–276
syntactical
  consistency, 52
  consistentcy, 55
  decidability, 54
  inconsistency, 54
syntactical consistency, 40
syntactical system, 846
syntax, 36, 37, 46
synthetic domain theory, 788–789
system
  simply infinite, 331
system of arithmetic, 528
Szabo, M., 672, 758
Szczerba, L. W., 572, 592, 593, 595, 608
Szmielew, W., 565, 575

$A^{\mathcal{T}}$, 209
$\mathbb{T}^{\omega}$ model, 783
$\mathcal{T}_n$, 209
T (Gödel's system), 489, 490, 493
T-sentence, 549, 553
T-sentences, 555
TACh_{λ}, 861
TACu, 866
TACu_{λ}, 862
tag, problem of, 625–628
Tait, W., 204, 218, 253, 325, 331, 683, 708, 766, 768

computability method, 753, 767, 768
confluence proof, 748
Takahashi, Masako, 723, 748
Takahashi, Moto-o, 768
Takahashi, S., 955
Takeuti, G., 378, 710, 712
Takeuti, G., system GLC, 768
Takeuti's conjecture, 768
Talcott, C., 997
Tammet, T., 776
Tanaka, K., 990, 991
Tapp, C., 325
Tarski's definition of truth, 541, 545, 547, 548, 552, 553, 559
Tarski's natural geometry, 596
Tarski's undefinability theorem, 551, 552
Tarski, A., 103, 199, 231, 253, 305, 307, 310, 312, 316–319, 386, 422, 438, 486, 511, 713, 736, 776, 902, 911–916, 923, 924, 929, 937, 941, 961, 966, 967, 973, 980, 988
(τ), *858*, 865
tautology, 30, 40, 49
sentential, 212
substitutional sentential, 212
Taylor, P., 789
Tennant, N., 995
Tennenbaum's Theorem, 474
Terese research group, 738
term, 845, *857*
combinatory, *864*
λ, *857*
Skolem, 210–211
term rewriting systems, 738
terminological explanations, 308, 311
terms, 46
terms as deductions, see *deductions as terms*735
The Cut Elimination Theorem, 402, 403
The Epsilon Substitution Method, 385, 389, 390, 407, 409, 430
The First Epsilon Theorem, 385, 393, 402, 406
The Law of the Excluded Middle, 397
The Paradox of Predication, 439, 440
The Second Epsilon Theorem, 386, 402, 405, 415
theorem, 547, 845
second recursion, 911
theorem proving
automated, 223, 233, 235, 240–242
human-oriented, 223, 233, 235
inductive, 233, 235
theorems, 40, 49, 50, 57
theoretical computer, 949
theoretical linguistics, 963
theory Q, 577
theory of atomistic Boolean algebras, 526
theory of beingles incomplete objects, 71
theory of Boolean algebras, 575
theory of definite descriptions, 1, 2, 22, 23
theory of dense linear orderings, 575
theory of dense orders, 526
theory of discrete orders, 526
theory of fields of sets, 526
theory of groups, 513
theory of identity, 526
theory of logical atomism, 1
theory of logical constructions, 1
theory of logical relations, 1
theory of mathematical knowledge, 17
theory of order, 513
theory of ordinals, 13
theory of real-closed fields, 567
theory of referece, 34
theory of relation arithmetic, 1, 14
theory of relations, 14
theory of the congruence of line segments, 527, 563
theory of the reals, 1, 13

theory of transfinite induction, 14
theory of types, 1, 10, 11, 24, 280, 287, 309
theory of vectors, 13
theory of well-orderings, 575
Thiel, C., 668, 715, 727
Thomas, W., 738
Thomason, R., 416, 417, 420, 422, 423, 428, 438
Thompson, F. B., 589
thought, 34
three-valued logic, 35
three-valued propositional semantics, 34, 41
Thue system, 175, 649–650
Thue, A., 175, 179, 649–650, 738
Tiles, M., 414, 441, 442
Tiuryn, J., 775
T_k, 969, 970
Toepell, M., 324, 329
topological interpretation, 103
topology
  κ-compact, 993
topos, elementary, 759
Torretti, R., 328
transfinite induction, 485, 667, 933, 935, 938
transfinite types, 737
translation
  $-_H$, *865*
  H to Λ, *865*
  $-_L$, *865*
  Λ to H, *865*
Troelstra, A. S., 81, 83, 84, 93, 96, 97, 103, 106, 111, 114–117, 120, 486, 493–495, 701, 757, 767
true predication, 66
truth, 530, 531, 536, 537, 541, 561, 564–567, 573, 607
  arithmetical, see *AT*875
  axiomatic theory of, 915, 916, 980, 983
  logics of, 968, 980
  self-referential, 943, 966, 968, 977, 978, 980, 983, 996–998
truth functionally complex proposition, 49
truth tables
  Bochvar-Kleene, 959, 960
  strong Kleene, 959, 960, 963
truth teller, 974
truth values, 38, 41, 42, 46, 49
truth-table, 620–623, 635
truth-value
  internal, 925
truth-value gaps, 961, 964
truth-value gluts, 971
truth-values space, see *space, of truth values*875
Turing Completeness Theorem, 475
Turing machine, 627–629, 631, 641–643, 649–650
Turing reducibility, 645, 647–648, 651–654
Turing, A. M., 490, 494
Turing, A., 379, 618, 619, 624, 632, 641–643, 647, 657, 732, 737, 823
  normalization proof, 766
  Turing computable, 732
Turner, D. A., 729
Turner, R., 967, 997
turnstile, 39, 40
Twardowski, K., 29, 305, 316, 318, 752
2, *860*
typable, 734, 774
  typability tests, 747, 774–776
type of all types, 764
type theory, 684, 903, 907, 908, 927, 929, 992
  constructive, 939, 940, 945
  Martin-Löf's, 939, 989
  ramified, 924
  simple, 912
type-free comprehension, 875
type-free foundations, 930, 931, 939

types, 737–738, 740–741, 747, 751–776, 824, 856, 860
approaches, four different, 751
atomic, 860
Curry's, 734, 740, 752
dependent, 740, 761, 764, 771
doctrine of, 884, 898, 899
Gentzen-style rules for, 740, 764
Martin-Löf's, 763–765
polymorphic, 771
principal, 774–776
recursive, 774
simple, 747, 774–776
Church's system, 733, 737, 753
theory of, 267
transfinite, 737
universe of, 939, 945
with intersection, 746, 772–774, 787
and ω, 773–774
types as propositions or formulae, see *propositions as types*734

ultraproduct, 977
ultraproducts, 133, 159, 565, 603
UMID, 955
undecidability, 566, 589, 911
undecidability of λ, 732, 744
undecidability of number theory, 473
undecidability of predicate logic, 473
undecidable, 575–577, 579, 580, 593, 594
undecidable theory, 473
undefinability, 911, 998
undefinedness
in λ, 779
in recursion theory, 780
undetermined, 41
nucleaer predications, 58
proposition, 49
truth value, 42
unification, 626
unification algorithm, 242–243, 775
uniform notation, see *notation, uniform*254
uniform substitution, 40
uniformity principle, 115
union, 41, 45
union axiom, 600
universal, 59, 69, 71, 544
quantification, 38
universal generality, 834
universal generators, 780
universal modified Kripke model, 118
universal spread, 111
universal Veldman model, 117
universality, 544, 559
universe of discourse, 41
unrestricted comprehension, 34
unrestricted freedom of assumption thesis, 46, 58
unsolvable problems, 624, 626–633, 648–650, 655, 657
unsolvable terms, 779–780
identification of, 780
unstable elements, 972
Urquhart, A., 891
Uspenskii, V., 777

Vafeiadou, G., 120
Vaihinger, H., 348
valid inference pattern, 49
validity, 49
Beth, 489
classical, 487
intuitionistic, 487
Kripke, 489
structural, 487–489
valuation, 960, 961, 973, 974
classical, 961, 971, 974
maximal, 962
partial, 960
PR-, 962
Schütte, 963
three-valued, 960
WK-, 962
valuations (partial, total), 768

van Dalen, D., 496
Van Eijck, J., 434
van Fraassen, B., 959–961, 963, 982
Van Lambalgen, M., 997
Vaporis, N., 120
variable, 268, *857*, 864
  bound, 857, *858*, 863
  bound *delta*-, 211
  bound γ-, 211
  free, *858*, 864
  free δ-, 211
  free γ-, 211
Variable binding term operators, 390, 404
variables, 38
Vaught, R. L., 173, 524, 526, 552, 564, 565, 608
Veblen, O., 485, 530, 671
Veblen, Oswald, 820, 826
Veldman's completeness proof, 116
Veldman's completeness theorem, 119
Veldman's intuitionistic completeness proof, 108
Veldman, W., 84, 108, 117, 489
Venn diagram, 48
Venn–Euler diagram, 42
Venneri, B., 773, 774
Venturini Zilli, M., 744
verifiability, 364
Vesley, R. E., 83, 110, 113, 114, 120
Vessiot, Ernest, 198
vicious circle principle, 891, 899, 903, 928
vicious circle principle (VCP), 10, 11
Vickers, S., 787
Vielheit, see *multiplicity*875
Vienna Circle, 264
Visser's rule, 101
Visser, A., 84, 101–103, 964, 966, 967, 972, 983
Vitali, G., 597
Volken, H., 749
von Neumann, J., 311, 354, 359, 365, 368, 369, 379, 465, 486, 487, 690, 903, 905
von Plato, J., 690, 694, 697, 702, 703, 706
von Schaper, H., 329
Vopenka, P., 989
de Vrijer, R. C., 749
Vuillemin, J., 750

W
  combinator, 917, 922–924
ω-consistency, 559
ω-incomplete theory, 559
ω-incompleteness, 559, 562
ω-rule, 562, 563
Waaldijk, F., 84
Wadsworth, C. P., 744, 749, 754, 779–780, 784
Wagon, S., 597
Wajsberg, M., 518, 521
Wand, M., 774, 784
Wang, H., 172, 227, 228, 231, 253, 254, 497–500
Warsaw, 529
Warsaw School of Logic, 518
Washington, G., 70
Wavre, R., 82
weak completeness proof, 57
weak completeness theorem, 54
weak counterexamples, 114
weak interpretability, 579
weak König lemma, 179
weak normalization, see *WN*766
weak reduction, 739
weakly inaccessible cardinal, 600
weakly interpretable, 579
Weaver, G., 49
Weber, H., 327
Wedderburn, Joseph H. M., 198, 252
Wegner, P., 774
Wehmaier, K., 992
Weiermann, A., 767
Weil, André, 197, 207
Weinstein, S., 84
Weirstrass, C., 337, 340, 341

Welch, P., 975, 976, 994
well-formed formula, 46, 846
well-formed formulas (wffs), 39, 41
well-ordering, 879, 882, 886–888, 896, 911, 945, 946, 984
Wells, J., 775
Wertverlauf operation, 919, 941
Weydert, E., 949
Weyl, E., 353
Weyl, H., 177, 180, 200, 204, 254, 348, 350, 359, 620, 630, 632, 895, 897, 898, 900–902
WF, 904
White, M., 607
White, R. B., 958, 995, 996
Whitehead, see also *Principia Mathematica*254
  Alfred North, 200
Whitehead, A. N., 2, 3, 12, 13, 15, 24, 87, 89, 135, 302, 307, 309, 311, 314, 323, 550, 600, 619–622, 625, 725, 728, 731
Whitman, P., 173
Wilken, G., 767
will, 278, 279
Williams, C. J. F., 436, 437
Wirth, Claus-Peter, 195, 211, 212, 216, 223, 224, 230, 235, 254
Wittgenstein, Ludwig, 199, 254, 498
WN, 766, 767, *see* normalization
Wolska, Magdalena, 244
Wood, A., 20
Woodruff, P., 926, 961, 962, 967
Woodward, P. M., 742
word problem
  for groups, 650
  for semigroups, 648–650
word problem for lattices, 173
word problem for semi-groups, 175
Wright, C., 442, 991

(ξ), *858*, 866

Y, see *fixed-point combinators*732
Yablo, S., 967, 987
Yanovskaya, S. A., 727
Yaqūb, 967
Yoneda's lemma, 952
Young, J. W., 530
Yourgrau, P., 491
Yuting, S., 500, 904

Zacchi, M., 787
Zach, R., 83, 90, 120, 176, 177, 325
Zaremba, Stanisław, 205
Zermelo's set theory, 601
Zermelo, E., 131, 134, 152, 176, 310, 340–342, 358, 499, 600, 601, 878, 879, 882, 886–888, 893, 895–897, 900–902, 907, 918, 928
Zermelo–Fraenkel set theory, 43
Zermelo-Fraenkel's axiom-system, 602
0, *860*
(ζ), 827, *859*
ZFC, 877, 946, 966, 996
zigzag theory, 890
Zilsel, E., 486, 491, 493
Zucker, J., 753
Zwicker, W. S., 988
$\{F\}(A)$, 918
$\Sigma^1_1 - AC$, 940, 984
$\rightarrow_i$, 942, 943
$\vdash_i$, 944
$\equiv$, 925, 952, 953, 970, 971, 991